

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:53:43 ; Search time 15.4667 seconds
(without alignments)
21.885 Million cell updates/sec

Title: US-09-606-129A-17
Perfect score: 28
Sequence: 1 QKXCCXXK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2.6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	75.0	115	3	US-09-401-869-5
2	21	75.0	166	4	US-09-252-991A-17918
3	21	75.0	273	4	US-09-252-991A-23086
4	21	75.0	338	2	US-08-933-750C-4
5	21	75.0	338	3	US-09-234-613-4
6	21	75.0	788	2	US-07-728-215-27
7	21	75.0	788	4	US-08-938-085A-27
8	21	75.0	788	4	US-10-072-844-27
9	21	75.0	922	2	US-08-464-402-2
10	21	75.0	922	3	US-09-054-775C-2
11	20	71.4	43	4	US-09-209-676-63
12	20	71.4	43	4	US-09-209-676-69
13	20	71.4	43	4	US-09-209-676-72
14	20	71.4	43	4	US-09-209-676-74
15	20	71.4	43	4	US-09-209-676-77
16	20	71.4	73	2	US-08-637-759B-406
17	20	71.4	73	3	US-08-871-355A-406
18	20	71.4	73	4	US-09-201-945-406
19	20	71.4	127	4	US-09-107-532A-7074
20	20	71.4	130	2	US-08-888-497-43
21	20	71.4	130	4	US-09-362-230-43
22	20	71.4	130	5	PCT-US94-07926-43
23	20	71.4	149	4	US-09-328-352-6466
24	20	71.4	158	2	US-08-888-497-22
25	20	71.4	158	4	US-09-362-230-22
26	20	71.4	158	5	PCT-US94-07926-22
27	20	71.4	174	4	US-09-134-001C-5091

28	71.4	232	4	US-09-216-393B-116	Sequence 116, Appl
29	71.4	245	1	US-07-876-284-2	Sequence 2, Appl1
30	71.4	245	1	US-08-276-151-9	Sequence 9, Appl1
31	71.4	246	1	US-08-276-151-7	Sequence 7, Appl1
32	71.4	252	2	US-08-685-992-33	Sequence 33, Appl
33	71.4	252	2	US-09-144-923-33	Sequence 33, Appl
34	71.4	280	4	US-09-134-001C-4524	Sequence 4524, Ap
35	71.4	331	1	US-08-356-180-3	Sequence 3, Appl1
36	71.4	496	2	US-08-838-543-2	Sequence 2, Appl1
37	71.4	577	2	US-07-728-215-29	Sequence 29, Appl
38	71.4	577	4	US-08-938-085A-29	Sequence 29, Appl
39	71.4	577	4	US-10-072-844-29	Sequence 29, Appl
40	71.4	584	1	US-08-161-290-1	Sequence 1, Appl1
41	71.4	584	2	US-08-450-755-1	Sequence 1, Appl1
42	71.4	585	1	US-08-117-907-2	Sequence 2, Appl1
43	71.4	585	1	US-08-485-718-11	Sequence 11, Appl
44	71.4	585	1	US-08-485-718-13	Sequence 13, Appl
45	71.4	585	2	US-08-484-530-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1
US-09-401-869-5
; Sequence 5, Application US/09401869
; Patent No. 6238901
; GENERAL INFORMATION:
; APPLICANT: Morgan, Richard D.
; APPLICANT: Xu, Qing
; TITLE OF INVENTION: A No. 6238901el Type II Restriction Endonuclease, Hpy188III,
; TITLE OF INVENTION: Obtainable From Helicobacter Pylori J188 And A Process
; TITLE OF INVENTION: For Producing The Same
; FILE REFERENCE: NEB-149
; CURRENT APPLICATION NUMBER: US/09/401,869
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-401-869-5

Query Match 75.0%; Score 21; DB 3; Length 115;
Best Local Similarity 50.0%; Pred. No. 5.8e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 QKXCCXXK 8
Db 42 QKACEVIK 49

RESULT 2
US-09-252-991A-17918
; Sequence 17918, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17918
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-17918

Query Match 75.0%; Score 21; DB 4; Length 166;
Best Local Similarity 50.0%; Pred. No. 7.7e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
Db 1 QKLCSRGK 8

RESULT 3

US-09-252-991A-23086
; Sequence 23086, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23086
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23086

Query Match 75.0%; Score 21; DB 4; Length 273;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
Db 110 QKGVHAK 117

RESULT 4

US-08-933-750C-4
; Sequence 4, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HMCINOT01
; CLONE: 9337
US-08-933-750C-4

Query Match 75.0%; Score 21; DB 2; Length 338;
Best Local Similarity 50.0%; Pred. No. 1.4e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
Db 155 QKSCERQK 162

RESULT 5

US-09-234-613-4
; Sequence 4, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

```
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HMCINOT01
; CLONE: 9337
; US-09-234-613-4

Query Match 75.0%; Score 21; DB 3; Length 338;
Best Local Similarity 50.0%; Pred. No. 1.4e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXX 8
Db 155 QKSCERQK 162

RESULT 6
US-07-728-215-27
; Sequence 27, Application US/07728215
; Patent No. 5962643
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; APPLICANT: Quaranta, Vito
; APPLICANT: Pytela, Robert
; TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/728,215
; FILING DATE: 19910711
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P31 8717
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-728-215-27

Query Match 75.0%; Score 21; DB 2; Length 788;
Best Local Similarity 50.0%; Pred. No. 2.7e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXX 8
Db 403 QKCSHMK 410

RESULT 7
US-08-938-085A-27
; Sequence 27, Application US/08938085A
; Patent No. 6339148
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; APPLICANT: Quaranta, Vito
; APPLICANT: Pytela, Robert
; TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,085A
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,215
; FILING DATE: 11-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parent, Annette S.
; REGISTRATION NUMBER: 42,058
; REFERENCE/DOCKET NUMBER: 023070-080210US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 788 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-938-085A-27

Query Match 75.0%; Score 21; DB 4; Length 788;
Best Local Similarity 50.0%; Pred. No. 2.7e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXX 8
Db 403 QKCSHMK 410

RESULT 8
US-10-072-844-27
; Sequence 27, Application US/10072844
; Patent No. 6576432
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; APPLICANT: Quaranta, Vito
; APPLICANT: Pytela, Robert
; TITLE OF INVENTION: A No. 6576432el Integrin Beta Subunit and Uses
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/072,844
FILING DATE: 06-Feb-2002
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/938,085A
FILING DATE: 26-SEP-1997
APPLICATION NUMBER: US/07/728,215
FILING DATE: 11-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Parent, Annette S.
REGISTRATION NUMBER: 42,058
REFERENCE/DOCKET NUMBER: 023070-080210US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 788 amino acids
TYPE: amino acid
-TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-10-072-844-27

Query Match 75.0%; Score 21; DB 4; Length 788;
Best Local Similarity 50.0%; Pred. No. 2.7e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
Db 403 QKCSHMK 410

RESULT 9
US-08-464-402-2
Sequence 2, Application US/08464402
Patent No. 5858705
GENERAL INFORMATION:
APPLICANT: WEI, ET AL.
TITLE OF INVENTION: Human DNA Ligase III
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,402
FILING DATE: June 5, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/03939
FILING DATE: 31 MAR 95
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-388
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 922 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-464-402-2

Query Match 75.0%; Score 21; DB 2; Length 922;
Best Local Similarity 50.0%; Pred. No. 3e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
Db 634 QKWCTVTK 641

RESULT 10
US-09-054-775C-2
Sequence 2, Application US/09054775C
Patent No. 6284504
GENERAL INFORMATION:
APPLICANT: Wei, Ying-Fei
Yu, Guo-Liang
Haseltine, William
TITLE OF INVENTION: Human DNA Ligase III
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/054,775C
FILING DATE: 03-Apr-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/464,402
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: PCT/US95/03939
FILING DATE: 31-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Hoover, Kenley K.
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PF161D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 922 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-054-775C-2

Query Match 75.0%; Score 21; DB 3; Length 922;
Best Local Similarity 50.0%; Pred. No. 3e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
Db 634 QKWCTVTK 641

```
RESULT 11
US-09-209-676-63
; Sequence 63, Application US/09209676
; Patent No. 6524856
; GENERAL INFORMATION:
; APPLICANT: Pangene Corporation
; TITLE OF INVENTION: The Use of Consensus Sequences for Targeted Homologous
; RECOMBINATION and Recombination in Gene Families
; FILE REFERENCE: A-65678-1/RFT/NBC
; CURRENT APPLICATION NUMBER: US/09/209,676
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/070,734
; PRIOR FILING DATE: 1997-12-11
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-676-63

Query Match          71.4%; Score 20; DB 4; Length 43;
Best Local Similarity 37.5%; Pred. No. 4.3e+02;
Matches 3; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
:| | |
Db 21 EKACSLAK 28

RESULT 12
US-09-209-676-69
; Sequence 69, Application US/09209676
; Patent No. 6524856
; GENERAL INFORMATION:
; APPLICANT: Pangene Corporation
; TITLE OF INVENTION: The Use of Consensus Sequences for Targeted Homologous
; RECOMBINATION and Recombination in Gene Families
; FILE REFERENCE: A-65678-1/RFT/NBC
; CURRENT APPLICATION NUMBER: US/09/209,676
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/070,734
; PRIOR FILING DATE: 1997-12-11
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Rat
US-09-209-676-69

Query Match          71.4%; Score 20; DB 4; Length 43;
Best Local Similarity 37.5%; Pred. No. 4.3e+02;
Matches 3; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
:| | |
Db 21 EKACSLAK 28

RESULT 13
US-09-209-676-72
; Sequence 72, Application US/09209676
; Patent No. 6524856
; GENERAL INFORMATION:
; APPLICANT: Pangene Corporation
; TITLE OF INVENTION: The Use of Consensus Sequences for Targeted Homologous
; RECOMBINATION and Recombination in Gene Families
; FILE REFERENCE: A-65678-1/RFT/NBC
; CURRENT APPLICATION NUMBER: US/09/209,676
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/070,734
; PRIOR FILING DATE: 1997-12-11
```

```
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Rat
US-09-209-676-72

Query Match          71.4%; Score 20; DB 4; Length 43;
Best Local Similarity 37.5%; Pred. No. 4.3e+02;
Matches 3; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
:| | |
Db 21 EKACSLAK 28

RESULT 14
US-09-209-676-74
; Sequence 74, Application US/09209676
; Patent No. 6524856
; GENERAL INFORMATION:
; APPLICANT: Pangene Corporation
; TITLE OF INVENTION: The Use of Consensus Sequences for Targeted Homologous
; RECOMBINATION and Recombination in Gene Families
; FILE REFERENCE: A-65678-1/RFT/NBC
; CURRENT APPLICATION NUMBER: US/09/209,676
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/070,734
; PRIOR FILING DATE: 1997-12-11
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 74
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Bovine
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-209-676-74

Query Match          71.4%; Score 20; DB 4; Length 43;
Best Local Similarity 37.5%; Pred. No. 4.3e+02;
Matches 3; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
:| | |
Db 21 EKACSLAK 28

RESULT 15
US-09-209-676-77
; Sequence 77, Application US/09209676
; Patent No. 6524856
; GENERAL INFORMATION:
; APPLICANT: Pangene Corporation
; TITLE OF INVENTION: The Use of Consensus Sequences for Targeted Homologous
; RECOMBINATION and Recombination in Gene Families
; FILE REFERENCE: A-65678-1/RFT/NBC
; CURRENT APPLICATION NUMBER: US/09/209,676
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/070,734
; PRIOR FILING DATE: 1997-12-11
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 77
; LENGTH: 43
; TYPE: PRT
; ORGANISM: Bovine
US-09-209-676-77

Query Match          71.4%; Score 20; DB 4; Length 43;
Best Local Similarity 37.5%; Pred. No. 4.3e+02;
Matches 3; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

Qy 1 QKXCXXX 8
: | |
Db 21 EKACSLAK 28

Search completed: July 29, 2003, 09:57:31
Job time : 16.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:55:03 ; Search time 19.2 Seconds
(without alignments)
49.483 Million cell updates/sec

Title: US-09-606-129a-17

Perfect score: 28
Sequence: 1 QKXCXXXK 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	23	82.1	182	9	US-09-780-717-14
2	22	78.6	74	10	US-09-796-692-1148
3	22	78.6	74	10	US-09-796-692-1296
4	22	78.6	74	10	US-09-796-692-2475
5	22	78.6	74	15	US-10-040-862-1148
6	22	78.6	74	15	US-10-040-862-1296
7	22	78.6	74	15	US-10-040-862-2475
8	22	78.6	121	15	US-10-106-698-4777
9	22	78.6	294	11	US-09-820-843A-7
10	22	78.6	203	9	US-09-815-242-5608
11	22	78.6	305	15	US-10-146-772-90
12	22	78.6	322	9	US-09-815-242-12459
13	22	78.6	617	10	US-09-764-864-1263
14	22	78.6	617	10	US-09-860-670-82
15	22	78.6	617	15	US-10-106-698-4649

16	21	75.0	79	9	US-09-925-299-1339	Sequence 1339, Ap
17	21	75.0	79	11	US-09-925-299-1339	Sequence 1339, Ap
18	21	75.0	104	15	US-10-082-830-211	Sequence 211, App
19	21	75.0	141	10	US-09-925-300-1686	Sequence 1686, Ap
20	21	75.0	170	15	US-10-153-668-404	Sequence 404, App
21	21	75.0	177	15	US-10-153-668-290	Sequence 290, App
22	21	75.0	188	10	US-09-764-864-859	Sequence 859, App
23	21	75.0	338	9	US-09-840-787-4	Sequence 4, Appli
24	21	75.0	454	9	US-09-815-242-11157	Sequence 11157, A
25	21	75.0	521	10	US-09-925-300-1667	Sequence 1667, Ap
26	21	75.0	541	9	US-09-815-242-11316	Sequence 11316, A
27	21	75.0	625	15	US-10-028-072-70	Sequence 70, Appl
28	21	75.0	625	15	US-10-121-049-70	Sequence 70, Appl
29	21	75.0	625	15	US-10-123-904-70	Sequence 70, Appl
30	21	75.0	625	15	US-10-140-470-70	Sequence 70, Appl
31	21	75.0	625	15	US-10-175-746-70	Sequence 70, Appl
32	21	75.0	625	15	US-10-176-918-70	Sequence 70, Appl
33	21	75.0	625	15	US-10-176-921-70	Sequence 70, Appl
34	21	75.0	625	15	US-10-137-865-70	Sequence 70, Appl
35	21	75.0	625	15	US-10-140-474-70	Sequence 70, Appl
36	21	75.0	625	15	US-10-142-431-70	Sequence 70, Appl
37	21	75.0	625	15	US-10-143-114-70	Sequence 70, Appl
38	21	75.0	625	15	US-10-140-002-70	Sequence 70, Appl
39	21	75.0	625	15	US-10-142-419-70	Sequence 70, Appl
40	21	75.0	625	15	US-10-123-262-70	Sequence 70, Appl
41	21	75.0	625	15	US-10-142-423-70	Sequence 70, Appl
42	21	75.0	625	15	US-10-121-050-70	Sequence 70, Appl
43	21	75.0	625	15	US-10-141-755-70	Sequence 70, Appl
44	21	75.0	625	15	US-10-143-032-70	Sequence 70, Appl
45	21	75.0	625	15	US-10-123-108-70	Sequence 70, Appl

ALIGNMENTS

RESULT 1

US-09-780-717-14
; Sequence 14, Application US/09780717
; Patent No. US20010044941A1
; GENERAL INFORMATION:
; APPLICANT: Helentjaris, Tim
; APPLICANT: Bates, Nic
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: NO. US20010044941A1 Invertase Inhibitors and Methods
; FILE OF INVENTION: of use
; FILE REFERENCE: 035718/208677
; CURRENT APPLICATION NUMBER: US/09/780,717
; PRIOR FILING DATE: 2001-02-09
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Vitis L
US-09-780-717-14

Query Match 82.1%; Score 23; DB 9; Length 182;
Best Local Similarity 50.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 QKXCXXXK 8

Db 34 QKCRSTK 41

RESULT 2

US-09-796-692-1148
; Sequence 1148, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander

```
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1148
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1148

Query Match      78.6%      Score 22; DB 10; Length 74;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
    |||
Db 38 QKMCSLSK 45

RESULT 3
US-09-796-692-1296
; Sequence 1296, Application US/09/796,692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1148
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-1148

Query Match      78.6%      Score 22; DB 10; Length 74;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
    |||
Db 38 QKMCSLSK 45

RESULT 4
US-09-796-692-2475
; Sequence 2475, Application US/09/796,692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2475
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-692-2475
```

Query Match 78.6%; Score 22; DB 10; Length 74;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 4; Gaps 0;

Oy 1 QKXCXXXK 8
Db 38 QKMC SLK 45

RESULT 5

US-10-040-862-1148
; Sequence 1148, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040.862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1148
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-040-862-1148

Query Match 78.6%; Score 22; DB 15; Length 74;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 4; Gaps 0;

Oy 1 QKXCXXXK 8
Db 38 QKMC SLK 45

RESULT 6

US-10-040-862-1296
; Sequence 1296, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther

;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: US 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: US 60/223,416
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: US 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; PRIOR APPLICATION NUMBER: US 09/796,692
;; PRIOR FILING DATE: 2001-03-01
;; NUMBER OF SEQ ID NOS: 10467
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 2475
;; LENGTH: 74
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-040-862-2475

Query Match 78.6%; Score 22; DB 15; Length 74;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
||| |
Db 38 QKMCSLSK 45

RESULT 8
US-10-106-698-4777
; Sequence 4777, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4777
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-4777

Query Match 78.6%; Score 22; DB 15; Length 121;
Best Local Similarity 50.0%; Pred. No. 4.4e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
||| |
Db 69 QKACTWAK 76

RESULT 9
US-09-820-843A-7
; Sequence 7, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 203
; TYPE: PRT
; ORGANISM: C. trachomatis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: histone like protein 2
; NAME/KEY: misc_feature
; OTHER INFORMATION: gi|3328438
US-09-820-843A-7

Query Match 78.6%; Score 22; DB 11; Length 203;
Best Local Similarity 50.0%; Pred. No. 6.7e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
||| |
Db 7 QKCKSTRK 14

RESULT 10
US-09-815-242-5608
; Sequence 5608, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5608
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Staphylococcus aureus

US-09-815-242-5608

Query Match 78.6%; Score 22; DB 9; Length 294;
Best Local Similarity 50.0%; Pred. No. 9e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
|||
Db 287 QKACQTLK 294

RESULT 11

US-10-146-772-90
; Sequence 90, Application US/10146772
; Publication No. US20030124698A1

; GENERAL INFORMATION:
; APPLICANT: Short, Jay

; APPLICANT: Weiner, David

; APPLICANT: Chaplin, Jennifer

; APPLICANT: Chi, Ellen

; APPLICANT: Milan, Aileen

; APPLICANT: DeSantis, Grace

; APPLICANT: Madden, Mark

; APPLICANT: Burk, Mark

; TITLE OF INVENTION: Nitrilases

; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US

; CURRENT APPLICATION NUMBER: US/10/146,772

; CURRENT FILING DATE: 2002-05-15

; PRIOR APPLICATION NUMBER: US 60/309,006

; PRIOR FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US 60/351,336

; PRIOR FILING DATE: 2002-01-22

; PRIOR APPLICATION NUMBER: US 60/300,189

; PRIOR FILING DATE: 2001-06-21

; PRIOR APPLICATION NUMBER: US 09/751,299

; PRIOR FILING DATE: 2000-12-28

; PRIOR APPLICATION NUMBER: US 60/254,414

; PRIOR FILING DATE: 2000-12-07

; PRIOR APPLICATION NUMBER: US 60/173,609

; PRIOR FILING DATE: 1999-12-29

; NUMBER OF SEQ ID NOS: 386

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 90

; LENGTH: 305

; TYPE: PRT

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: Obtained from an environmental sample

US-10-146-772-90

Query Match 78.6%; Score 22; DB 15; Length 305;

Best Local Similarity 50.0%; Pred. No. 9.3e+02;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
|||
Db 87 QKLCQAAK 94

RESULT 12

US-09-815-242-12459

; Sequence 12459, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12459
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12459

Query Match 78.6%; Score 22; DB 9; Length 322;
Best Local Similarity 50.0%; Pred. No. 9.7e+02;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
|||
Db 301 QKACQTLK 308

RESULT 13

US-09-764-864-1263

; Sequence 1263, Application US/09764864

; Patent No. US20020132753A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT23

; CURRENT APPLICATION NUMBER: US/09/764,864

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 1792

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1263

; LENGTH: 617

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-864-1263

Query Match 78.6%; Score 22; DB 10; Length 617;

Best Local Similarity 50.0%; Pred. No. 1.6e+03;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
|||
Db 529 QKDCLASK 536

RESULT 14

US-09-860-670-82

; Sequence 82, Application US/09860670

; Patent No. US20020165137A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA127P1

; CURRENT APPLICATION NUMBER: US/09/860,670

; CURRENT FILING DATE: 2001-05-21

; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-860-670-82 .

Query Match 78.6%; Score 22; DB 10; Length 617;
Best Local Similarity 50.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
Db 529 QKDCLASK 536

RESULT 15
US-10-106-698-4649
; Sequence 4649, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4649
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-4649

Query Match 78.6%; Score 22; DB 15; Length 617;
Best Local Similarity 50.0%; Pred. No. 1.6e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
Db 529 QKDCLASK 536

Search completed: July 29, 2003, 09:58:14
Job time : 20.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:53:07 ; Search time 13.8667 seconds
(without alignments)
55.482 Million cell updates/sec

Title: US-09-606-129A-17
Perfect score: 28
Sequence: 1 QKXCXXXK 8
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	78.6	101	2 T09149	late-embryogenesis
2	22	78.6	149	2 A81302	homolog of BLC pro
3	22	78.6	203	2 D71563	probable histone-1
4	22	78.6	322	2 G89784	conserved hypotet
5	22	78.6	419	2 H96601	hypothetical prote
6	22	78.6	451	2 E97234	ethanolamine ammon
7	22	78.6	642	2 T33262	hypothetical prote
8	22	78.6	683	2 T03146	probable glycoprot
9	22	78.6	702	2 T03903	hypothetical prote
10	22	78.6	858	2 S68227	ubiquitin thiolest
11	22	78.6	870	2 T09137	extensin homolog -
12	22	78.6	946	2 A71843	d-lactate dehydrog
13	22	78.6	3507	2 T34513	hypothetical prote
14	21	75.0	82	2 D64699	hypothetical prote
15	21	75.0	86	2 G71819	hypothetical prote
16	21	75.0	149	2 A81026	hypothetical prote
17	21	75.0	170	2 I48297	hypothetical prote
18	21	75.0	170	2 A55499	cellular nucleic a
19	21	75.0	171	2 I48298	zinc finger protei
20	21	75.0	177	2 A32760	cellular nucleic a
21	21	75.0	177	2 J23512	cellular nucleic a
22	21	75.0	178	2 I49259	cellular nucleic a
23	21	75.0	192	2 JS0373	cellular nucleic a
24	21	75.0	210	2 D81311	thiamine-phosphate
25	21	75.0	247	2 B82730	uridylylate kinase X
26	21	75.0	262	2 T22371	hypothetical prote
27	21	75.0	275	2 H71690	hypothetical prote
28	21	75.0	276	2 S20690	31.6K hypothetical
29	21	75.0	304	2 S69034	probable membrane

30 21 75.0 305 2 C97169 uncharacterized pr
31 21 75.0 309 2 H71328 probable flagellar
32 21 75.0 320 2 A84708 hypothetical prote
33 21 75.0 320 2 H71860 probable type II D
34 21 75.0 323 2 H90119 hypothetical prote
35 21 75.0 333 2 T05984 hypothetical prote
36 21 75.0 344 2 T34421 hypothetical prote
37 21 75.0 350 2 S57062 hypothetical prote
38 21 75.0 360 2 G72235 hypothetical prote
39 21 75.0 415 2 T12974 hypothetical prote
40 21 75.0 417 2 A72236 hypothetical prote
41 21 75.0 439 2 T30328 hsdS protein - Kie
42 21 75.0 454 2 C64104 gamma-aminobutyrat
43 21 75.0 471 2 E96573 protein F12M16.23
44 21 75.0 491 2 C90092 hypothetical prote
45 21 75.0 492 2 S18445 variant surface gl

ALIGNMENTS

RESULT 1
T09149
late-embryogenesis abundant protein EMB16 - white spruce (fragment)
C:Species: Picea glauca (white spruce)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T09149
R:Dong, J.Z.; Dunstan, D.I.
submitted to the EMBL Data Library, June 1996
A:Description: Gene expression during somatic embryogenesis.
A:Reference number: Z16588
A:Accession: T09149
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-101 <DON>
A:Cross-references: EMBL:L47629; NID:gl350498; PID:gl350499
C:Genetics:
A:Gene: EMB16

Query Match 78.6%; Score 22; DB 2; Length 101;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
Db 8 QKSCSDQK 15

RESULT 2

A81302
homolog of BLC protein Cj1547 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: A81302
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Bar
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: A81302
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-149 <PAR>
A:Cross-references: GB:AL1139078; GB:AL111168; NID:g668723; PIDN:CAB73963.1; PID:g696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1547

Query Match 78.6%; Score 22; DB 2; Length 149;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8

Db 28 QKSCLSNK 35

RESULT 3

D71563

probable histone-like protein 2 - Chlamydia trachomatis

C:Species: Chlamydia trachomatis

C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 21-Jul-2000

C:Accession: D71563; JN0852

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis

A:Reference number: AV1570; MUID:99000809; PMID:9784136

A:Accession: D71563

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-203 <ARN>

A:Cross-references: GB:AE001279; GB:AE001273; NID:g3328434; PIDN:AAC67637.1; PID:g332843

A:Experimental source: serotype D, strain UW-3/Cx

R:Hackstadt, T.; Brickman, T.J.; Barry III, C.E.; Sager, J. Gene 132, 137-141, 1993

A:Title: Diversity in the Chlamydia trachomatis histone homologue Hc2.

A:Reference number: JN0851; MUID:94010337; PMID:8406036

A:Accession: JN0852

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 3-203 <HAC>

A:Cross-references: GB:LI2963; NID:g289838; PIDN:AAA23131.1; PID:g289839

C:Genetics:

A:Gene: hctB

C:Superfamily: histone H1

C:Keywords: chromosomal protein

Query Match 78.6%; Score 22; DB 2; Length 203;
Best Local Similarity 50.0%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
|||

Db 7 QKCKSTRK 14

RESULT 4

G89784

conserved hypothetical protein SA0212 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: G89784

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K. Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: G89758; MUID:21311952; PMID:11418146

A:Accession: G89784

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-322 <KUR>

A:Cross-references: GB:BA000018; PID:g13700135; PIDN:BA641434.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA0212

Query Match 78.6%; Score 22; DB 2; Length 322;
Best Local Similarity 50.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
|||

Db 301 QKACQTLK 308

RESULT 5

H96601

hypothetical protein T6H22.14 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: H96601

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L. Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H96601

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-419 <STO>

A:Cross-references: GB:AE005173; NID:g6056379; PIDN:AAF02843.1; GSPDB:GN00141

C:Genetics:

A:Gene: T6H22.14

A:Map position: 1

C:Superfamily: yeast probable purine nucleotide-binding protein YBR025C

Query Match 78.6%; Score 22; DB 2; Length 419;
Best Local Similarity 50.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
|||

Db 42 QKRCFSSK 49

RESULT 6

E97234

ethanolamine ammonia lyase large chain [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: E97234

R:Noelling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R. J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: E97234

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-451 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80664.1; PID:g15025752; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum AFCC824

C:Genetics:

A:Gene: CAC2718

Query Match 78.6%; Score 22; DB 2; Length 451;
Best Local Similarity 50.0%; Pred. No. 4.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
|||

Db 137 QKICNTAK 144

RESULT 7

T33262

hypothetical protein C24B9.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33262

R:Murray, J.; Wohldmann, P.; Langston, Y.; O'Neal, D. submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid C24B9.

A:Reference number: Z21310

A:Accession: T33262

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-642 <MUR>

A:Cross-references: EMBL:AF068709; PIDN:AA019255.1; GSPDB:GN00023; CESP:C24B9.7

A:Experimental source: strain Bristol N2; clone C24B9

C:Genetics:

A:Gene: CESP:C24B9.7

A:Map position: 5

A:Introns: 63/3; 135/3; 294/3; 320/3; 388/3; 460/3; 542/1; 602/3

Query Match

Best Local Similarity 78.6%; Score 22; DB 2; Length 642;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8

DB 222 QKNCSTK 229

RESULT 8

T03146

probable glycoprotein A8 - alcelaphine herpesvirus 1

C:Species: alcelaphine herpesvirus 1

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999

C:Accession: T03146

R:Ensser, A.; Pfanz, R.; Fleckenstein, B.

J. Virol. 71, 6517-6525, 1997

A:Title: Primary structure of the alcelaphine herpesvirus 1 genome.

A:Reference number: Z14840; MUID:97404659; PMID:9261371

A:Accession: T03146

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-683 <ENS>

A:Cross-references: EMBL:AF005370; NID:92337967; PIDN:AA058098.1; PID:92338014

Query Match

Best Local Similarity 78.6%; Score 22; DB 2; Length 683;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8

DB 254 QKNCSTK 261

RESULT 9

T03903

hypothetical protein F39G3.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 17-Nov-2000

C:Accession: T03903

R:Du, Z.; Le, T.T.; Holmes, A.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid F39G3.

A:Reference number: Z15131

A:Accession: T03903

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-702 <DUZ>

A:Cross-references: EMBL:AF016424; NID:92291203; PIDN:AA065331.1; PID:92291209

C:Genetics:

A:Map position: V

A:Introns: 35/2; 78/2; 135/3; 299/3; 313/1; 332/1; 376/1; 476/3; 564/1; 599/2; 627/2

A:Note: F39G3.6

C:Superfamily: Caenorhabditis elegans hypothetical protein T26H2.7

Query Match

Best Local Similarity 78.6%; Score 22; DB 2; Length 702;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8

Db 247 QKFCVSK 254

RESULT 10

S68227

ubiquitin thiolesterase (EC 3.1.2.15) - human

N:Alternate names: ubiquitin isopeptidase T

C:Species: Homo sapiens (man)

C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999

C:Accession: S68227; S74295; G02070

R:Faquet, L.; Paquet, N.; Frutiger, S.; Hughes, G.J.; Hoang-Van, K.; Jaton, J.C.

FEBS Lett. 376, 233-237, 1995

A:Title: cDNA cloning of a human 100 kDa de-ubiquitinating enzyme: the 100 kDa human

A:Reference number: S68227; MUID:96105388; PMID:7498549

A:Accession: S68227

A:Molecule type: mRNA

A:Residues: 1-858 <FAL>

A:Cross-references: EMBL:X91349; NID:91122277; PIDN:CAA62690.1; PID:e208113; PID:g112

A:Accession: S74295

A:Molecule type: protein

A:Residues: 48-67; 104-118; 163-173; 205-218; 239-247; 292-317; 353-359; 417-429; 447-461; 483

R:Wilkinson, K.

submitted to the EMBL Data Library, August 1995

A:Reference number: G09157

A:Accession: G02070

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2, 'EL', 5-44, 'I', 46-467, 'R', 469-628, 'A', 653-680, 'D', 682-858 <WIL>

A:Cross-references: EMBL:U35116; NID:91008541; PIDN:AAA78934.1; PID:g1008542

C:Keywords: thiolester hydrolase

Query Match

Best Local Similarity 78.6%; Score 22; DB 2; Length 858;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8

DB 835 QKVCASEK 842

RESULT 11

T09137

extensin homolog - Trypanosoma brucei (fragment)

C:Species: Trypanosoma brucei

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: T09137

R:Bringaud, F.; Vedrenne, C.; Cuivillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, M.

Mol. Biochem. Parasitol. 94, 249-264, 1998

A:Title: Conserved organization of genes in trypanosomatids.

A:Reference number: Z16580; MUID:98418771; PMID:9747975

A:Accession: T09137

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-870 <BRI>

A:Cross-references: EMBL:AF031926; NID:g3452211; PIDN:AA032775.1; PID:g3452216

A:Experimental source: strain Antat1

Query Match

Best Local Similarity 78.6%; Score 22; DB 2; Length 870;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8

DB 95 QKRCSSK 102

RESULT 12

A71843

d-lactate dehydrogenase - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: A71843
R:Alm, R.A.; Ling, R.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: A71843
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-946 <ARN>
A:Cross-references: GB:AE001542; GB:AE001439; NID:94155739; PIDN:AAD06729.1; PID:9415575
A:Experimental source: strain J99
C:Genetics:
A:Gene: dld

Query Match 78.6%; Score 22; DB 2; Length 946;

Best Local Similarity 50.0%; Pred. No. 7.7e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
||| |
Db 54 QKLCALAK 61

RESULT 13

T34513

hypothetical protein ZK783.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34513

R:Favella, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994

A:Description: The sequence of C. elegans cosmid ZK783.

A:Reference number: 221536

A:Accession: T34513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3507 <FAB>

A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GNO0021; CESP:ZK783.1

A:Experimental source: strain Bristol N2; clone ZK783

C:Genetics:

A:Gene: CESP:ZK783.1

A:Map position: 3

A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Query Match

78.6%; Score 22; DB 2; Length 3507;

Best Local Similarity 50.0%; Pred. No. 2.2e+03;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
||| |
Db 3141 QKSCRTDK 3148

RESULT 14

D64699

hypothetical protein HP1436 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C:Accession: D64699

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKen

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: D64699

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-82 <TOM>

A:Cross-references: GB:AE000644; GB:AE000511; NID:92314609; PIDN:AAD08488.1; PID:9231462

Query Match 75.0%; Score 21; DB 2; Length 82;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
||| |
Db 66 QKACLLHK 73

RESULT 15

G71819

hypothetical protein jhp1329 - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: G71819

R:Alm, R.A.; Ling, R.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: G71819

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-86 <ARN>

A:Cross-references: GB:AE001556; GB:AE001439; NID:94155938; PIDN:AAD06905.1; PID:9415

A:Experimental source: strain J99

C:Genetics:

A:Gene: jhp1329

Query Match

75.0%; Score 21; DB 2; Length 86;

Best Local Similarity 50.0%; Pred. No. 2.1e+02;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
||| |
Db 70 QKACLLHK 77

Search completed: July 29, 2003, 09:56:59

Job time : 16.8667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:52:17 ; Search time 8 Seconds
(without alignments)
47.027 Million cell updates/sec

Title: US-09-606-129A-17
Perfect score: 28
Sequence: 1 QKXCXXXK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	22	78.6	201	1	HC2D_CHLTR	006280	chlamydia t
2	22	78.6	277	1	1433_EIMTE	098436	eimeria ten
3	22	78.6	858	1	UBP5_HUMAN	P45974	homo sapien
4	22	78.6	858	1	UBP5_MOUSE	P56399	mus musculus
5	21	75.0	133	1	IAMY_COILA	P15326	coix lachry
6	21	75.0	158	1	Y427_BUCAP	Q889c1	buchnera ap
7	21	75.0	170	1	CNBP_MOUSE	P53996	mus musculus
8	21	75.0	172	1	CNBP_CHICK	O42395	gallus gall
9	21	75.0	177	1	CNBP_HUMAN	P20694	homo sapien
10	21	75.0	192	1	Y3_SOCMV	P15633	soybean chl
11	21	75.0	210	1	THIE_CAMJE	Q9pn13	campylobact
12	21	75.0	247	1	PYRH_XYLFA	Q9beh0	xyella fas
13	21	75.0	309	1	FLIH_TREPA	O83416	treponema p
14	21	75.0	350	1	DPD3_YEAST	P47110	saccharomyc
15	21	75.0	362	1	NK23_MOUSE	P97334	mus musculus
16	21	75.0	364	1	NR23_HUMAN	Q8tau0	homo sapien
17	21	75.0	454	1	DAT_HAEIN	P44951	h diaminobu
18	21	75.0	492	1	VSM1_TRYBB	P26331	trypanosoma
19	21	75.0	541	1	SYR_HELPJ	Q9zmb9	helicobacte
20	21	75.0	541	1	SYR_HELPJ	P56128	helicobacte
21	21	75.0	632	1	FMN2_HUMAN	Q9nz56	homo sapien
22	21	75.0	653	1	FMN2_HUMAN	Q969c6	homo sapien
23	21	75.0	668	1	MTMW_METWO	O59647	methanobact
24	21	75.0	716	1	DING_ECOLI	Q09219	escherichia
25	21	75.0	753	1	YF6A_CABEL	Q09219	caenorhabdi
26	21	75.0	787	1	ITB6_MOUSE	Q920t9	mus musculus
27	21	75.0	788	1	ITB6_HUMAN	P18564	homo sapien
28	21	75.0	819	1	ZF95_MOUSE	Q9z1d8	mus musculus
29	21	75.0	922	1	DNL3_HUMAN	P49916	homo sapien
30	21	75.0	1015	1	DNL3_MOUSE	P97386	mus musculus
31	21	75.0	1234	1	JMJ_MOUSE	Q62315	mus musculus
32	21	75.0	1266	1	JMJ_HUMAN	Q92833	homo sapien
33	21	75.0	1567	1	FMN2_MOUSE	Q9j104	mus musculus

ALIGNMENTS

RESULT 1

ID	HC2D_CHLTR	STANDARD	PRT	201 AA
AC	Q06280			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Histone-like protein HC2.			
GN	HCTB OR CT046.			
OS	Chlamydia trachomatis.			
OC	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxID=813;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=D/UN-3/Cx;			
RA	MEDLINE=94010337; PubMed=8406036;			
RX	Hackstadt T., Brickman T.J., Barry C.E. III, Sager J.;			
RT	"Diversity in the Chlamydia trachomatis histone homologue Hc2.";			
RL	Gene 132:137-141(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=D/UN-3/Cx;			
RA	MEDLINE=99000809; PubMed=9784136;			
RX	Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;			
RT	"Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";			
RL	Science 282:754-759(1998).			
CC	-!- FUNCTION: MIGHT HAVE A ROLE IN ESTABLISHING THE NUCLEOID STRUCTURE OF ELEMENTARY BODIES.			
CC	-!- DEVELOPMENTAL STAGE: SPECIFIC TO THE EB (ELEMENTARY BODY) FORM IN THE LIFE CYCLE OF CHLAMYDIAE.			
CC	-!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY. HCT SUBFAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; L12963; AAA23131.1; .			
DR	EMBL; AE001279; AAC67637.1; .			
DR	PIR; D71563; D71563.			
KW	DNA-binding; Complete proteome.			
SQ	SEQUENCE 201 AA; 21403 MW; 4077EE90050BD079 CRC64;			

Query Match 78.6%; Score 22; DB 1; Length 201;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 QKXCXXXK 8
DB 5 QKXCSTRK 12

```

RESULT 2
1433_EIMTE
ID 1433_EIMTE STANDARD; PRT; 277 AA.
AC O96436;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 14-3-3 protein.
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
OX NCBI_TaxID=5802;
RN [1]_TaxID=5802;
SEQUENCE FROM N.A.
RC STRAIN=LS18;
RA Myers R.W., Liberator P.A., Allocco J.J., Anderson J.W., Sardana M.K.,
RA Wood T.L., Griffin P.R., Fujioka H., Schmatz D.M.;
RT "14-3-3 protein regulation of protozoan mannitol metabolism via
RT inhibition of mannitol-1-phosphate dehydrogenase.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the 14-3-3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF055715; AAD02687.1; -
CC HSP; P29312; I438.
CC InterPro; IPR000308; 14-3-3.
CC Pfam; PF00244; 14-3-3; 1.
CC PRINTS; PR00305; 1433ZETA.
CC SMART; SM00101; 14_3_3; 1.
CC PROSITE; PS00796; 1433_1; 1.
CC PROSITE; PS00797; 1433_2; 1.
CC SEQUENCE 277 AA; 31660 MW; 749A6D6A07267D5 CRC64;
-----
Query Match 78.6%; Score 22; DB 1; Length 277;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 QKXXXXX 8
Db 206 QKACEMAK 213
-----
RESULT 3
UBP5_HUMAN
ID UB5_HUMAN STANDARD; PRT; 858 AA.
AC P45974;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 5 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 5) (Ubiquitin-specific processing protease 5)
DE (Deubiquitinating enzyme 5) (Isopeptidase T).
DE USP5 OR ISOT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP MEDLINE=96105388; PubMed=7498549;
RA Falquet L., Paquet N., Frutiger S., Hughes G.J., Hoang-Van K.,
RA Jaton J.-C.;
RT "cDNA cloning of a human 100 kDa de-ubiquitinating enzyme: the 100
RT kDa human de-ubiquitinase belongs to the ubiquitin C-terminal

```

```

RT hydrolase family 2 (UCH2).";
RL FEBS Lett. 376:233-237(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96303695; PubMed=8723724;
RA Ansari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,
RA Malley T., Gibbs R.A.;
RT "A gene-rich cluster between the CD4 and triosephosphate isomerase
RT genes at human chromosome 12p13.";
RL Genome Res. 6:314-326(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Tashayev V.L., O'Connor L.B., Larsen C.N., Kasperek E., Pickart C.M.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=23388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Shevchenko Y., Bouffard G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=95154450; PubMed=7851534;
RA Falquet L., Paquet N., Frutiger S., Hughes G.J., Hoang-Van K.,
RA Jaton J.-C.;
RT "A human de-ubiquitinating enzyme with both isopeptidase and
RT peptidase activities in vitro.";
RL FEBS Lett. 359:73-77(1995).
CC -!- FUNCTION: CLEAVES LINEAR AND BRANCHED MULTIUBIQUITIN POLYMERS WITH
CC A MARKED PREFERENCE FOR BRANCHED POLYMERS.
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- COFACTOR: Zinc.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P45974-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P45974-2; Sequence=VSP_005259;
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC -!- SIMILARITY: Contains 2 UBA domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X91349; CAA62690.1; -
CC EMBL; U47927; AAC50465.1; -
CC EMBL; U47924; AAB51314.1; -
CC EMBL; U47924; AAB51315.1; -

```

```

DR EMBL; U35116; AAA78934.1;
DR EMBL; BC005139; AAH05139.1;
DR PIR; S68227; S68227.
DR MEROPS; C19.001;
DR GenBank; HGNC:12628; USP5.
DR MIM; 601447;
DR GO; GO:0004212; F:lysosomal cysteine-type endopeptidase; TAS.
DR GO; GO:0008581; F:ubiquitin isopeptidase T activity; TAS.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR001394; UCH-2.
DR InterPro; IPR001607; Znf_UBP.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00443; UCH; 1.
DR Pfam; PF02148; zf-UBP; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00290; Znf_UBP; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00973; UCH_2_3; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00973; UCH_2_3; 1.
KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family;
KW Alternative splicing; Zinc; Repeat.
FT DOMAIN 654 695 UBA 1.
FT DOMAIN 722 762 UBA 2.
FT ACT_SITE 335 335 BY SIMILARITY.
FT ACT_SITE 809 809 BY SIMILARITY.
FT ACT_SITE 818 818 BY SIMILARITY.
FT VARSPIC 629 652 GSLGFGYGNEDSFCSPHFSPTS -> A (in isoform Short)
FT /FTID-VSP_005259.
FT EL -> DV (IN REF. 1).
FT EL -> V (IN REF. 1).
FT K -> R (IN REF. 3).
FT G -> D (IN REF. 3).
SQ SEQUENCE 858 AA; 95786 MW; E99CB/CDFAG682C65 CRC64;

Query Match 78.6%; Score 22; DB 1; Length 858;
Best Local Similarity 50.0%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
DB 835 QKVCASEK 842

RESULT 4
UBP5_MOUSE STANDARD; PRT; 858 AA.
AC UPB5_MOUSE
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Ubiquitin carboxyl-terminal hydrolase 5 (EC 3.1.2.15) (Ubiquitin thiolesterase 5) (Ubiquitin-specific processing protease 5) (Deubiquitinating enzyme 5) (Isopeptidase 7).
GN USP5 OR ISOT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Anseri-Lari M.A., Oeltjen J.C., Schwartz S., Zhang Z., Muzny D.M., Lu J., Gorrell J.H., Chinault A.C., Belmont J.W., Miller W., Gibbs R.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CLEAVES LINEAR AND BRANCHED MULTIBIQUITIN POLYMERS WITH A MARKED PREFERENCE FOR BRANCHED POLYMERS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O = ubiquitin + a thiol.
CC -!- COFACTOR: ZINC.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC -!- SIMILARITY: Contains 2 UBA domains.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC002397; AAC36015.1;
DR MEROPS; C19.001;
DR MGD; MGI:1347343; Usp5.
DR InterPro; IPR000449; UBA_domain.
DR InterPro; IPR001394; UCH-2.
DR InterPro; IPR001607; Znf_UBP.
DR Pfam; PF00627; UBA; 2.
DR Pfam; PF00443; UCH; 1.
DR Pfam; PF02148; zf-UBP; 1.
DR SMART; SM00165; UBA; 2.
DR SMART; SM00290; Znf_UBP; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00973; UCH_2_3; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS00973; UCH_2_3; 1.
KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family;
KW Zinc; Repeat.
FT DOMAIN 654 695 UBA 1.
FT DOMAIN 722 762 UBA 2.
FT ACT_SITE 335 335 BY SIMILARITY.
FT ACT_SITE 809 809 BY SIMILARITY.
FT ACT_SITE 818 818 BY SIMILARITY.
SQ SEQUENCE 858 AA; 95833 MW; EIF31AOC92687CA8 CRC64;

Query Match 78.6%; Score 22; DB 1; Length 858;
Best Local Similarity 50.0%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
DB 835 QKVCASEK 842

RESULT 5
IAMV_COILA STANDARD; PRT; 133 AA.
AC P15326.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Alpha-amylase inhibitor/endochitinase (EC 3.2.1.14) (Fragments).
OS Coix lachryma-jobi (Jobs'tears).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Coix.
OX NCBI_TaxID=4505;
RN [1]
RP SEQUENCE.
RC TISSUE=Seed;
RX MEDLINE=90105532; PubMed=2605263;
RA Ary M.B., Richardson M., Shewry P.R.;
RT "Purification and characterization of an insect alpha-amylase inhibitor/endochitinase from seeds of Job's Tears (Coix lachryma-jobi).";
RL Biochim. Biophys. Acta 999:260-266(1989).
CC -!- FUNCTION: THIS PROTEIN FUNCTIONS BOTH AS AN ALPHA-AMYLASE INHIBITOR AND AS A CHITINASE.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-acetyl-D-glucosamine polymers of chitin.
CC -!- SIMILARITY: BELONGS TO CHITINASE CLASS IA WHICH IS SIMILAR TO CLASS IB IN THE CATALYTIC PART BUT WHICH INCLUDE A N-TERMINAL CHITIN-BINDING LECTIN DOMAIN (BELONGS TO FAMILY 19 OF GLYCOSYL HYDROLASES).
DR HSSP; P23951; 2BAA.

```

DR InterPro; IPR001002; Chitin_binding_1.
 DR InterPro; IPR000726; Glyco_hydro.19.
 DR ProDom; PD354900; Glyco_hydro.19.
 DR PROSITE; PS00026; CHITIN_BINDING; PARTIAL.
 DR PROSITE; PS00773; CHITINASE_19_1; PARTIAL.
 DR PROSITE; PS00774; CHITINASE_19_2; PARTIAL.
 KW Alpha-amylase inhibitor; Hydrolase; Glycosidase; Chitin degradation;
 Chitin-binding.
 FT NON_TER 1 1
 FT NON_CONS 14 15
 FT VARIANT 21 21 L -> I.
 FT NON_CONS 23 24
 FT NON_CONS 32 33
 FT VARIANT 89 89 L -> I.
 FT VARIANT 99 99 L -> I.
 FT NON_CONS 108 109
 FT NON_CONS 114 115
 FT NON_TER 133 133
 SQ SEQUENCE 133 AA; 14303 MW; E7E02ED7041B4F5E CRC64;
 Query Match 75.0%; Score 21; DB 1; Length 133;
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 QKXCXXXK 8
 DB 43 QKPCAAGK 50
 RESULT 6
 Y427_BUCAP STANDARD; PRT; 158 AA.
 ID Y427_BUCAP
 AC Q8K9C1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical UPF0054 protein BUS9427.
 GN BUS9427.
 OS Buchnera aphidicola (subsp. Schizaphis graminum).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=98794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tanas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
 RA Wernegreen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;
 RL "50 million years of genomic stasis in endosymbiotic bacteria."; Science 296:2376-2379(2002).
 CC -!- SIMILARITY: Belongs to the UPF0054 family.
 CC -!- CAUTION: This is a conceptual translation; a potential frameshift was corrected in position 40.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE014118; AAM67970.1; ALT_FRAME.
 CC HAMAP; MF00009; 1.
 CC InterPro; IPR002036; UPF0054.
 CC Pfam; PF02130; UPF0054; 1.
 CC ProDom; PD005688; UPF0054; 1.
 CC TIGRFAMs; TIGR00043; TIGR00043; 1.
 CC PROSITE; PS01306; UPF0054; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 158 AA; 19109 MW; 5817EA87E15FB31E CRC64;
 Query Match 75.0%; Score 21; DB 1; Length 158;
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 QKXCXXXK 8
 DB 12 QKNCKNTR 19
 RESULT 7
 CNBP_MOUSE STANDARD; PRT; 170 AA.
 ID CNBP_MOUSE
 AC P53996;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cellular nucleic acid binding protein (CNBP).
 GN ZNF9 OR CNBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Warden C.H., Krisans S.K., Purcell-Huynh D., Leete L.M.,
 RA Daluiski A., Diep A., Taylor B.A., Lusis A.J.;
 RT "Mouse cellular nucleic acid binding proteins: a highly conserved family identified by genetic mapping and sequencing."; Genomics 24:14-19(1994).
 RL Genomics 24:14-19(1994).
 CC -!- FUNCTION: SINGLE STRANDED DNA-BINDING PROTEIN, WITH SPECIFICITY TO THE STEROL REGULATORY ELEMENT (SRE). CNBP IS INVOLVED IN STEROL-MEDIATED REPRESSION.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ALSO PRESENT IN ENDOPLASMIC RETICULUM.
 CC -!- TISSUE SPECIFICITY: PRESENT IN ALL TISSUES EXAMINED.
 CC -!- SIMILARITY: TO S.POMBE BYR3 AND TO RETROVIRAL NUCLEIC ACID BINDING PROTEINS (NBP).
 CC
 CC -!- SIMILARITY: Contains 7 CCHC-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L12693; AAA89198.1;
 CC EMBL; Z11870; CAA77896.1;
 CC EMBL; X63866; CAA45345.1;
 CC EMBL; Z11871; CAA77897.1;
 CC EMBL; U20326; AAB60490.1;
 CC PIR; I48297; I48297.
 CC PIR; I48298; I48298.
 CC MGD; MGI:88431; Cnbp.
 CC GO; GO:0005829; C:cytosol; IDA.
 CC GO; GO:0005783; C:endoplasmic reticulum; IDA.
 CC InterPro; IPR001878; Znf_CCHC.
 CC Pfam; PF00098; zf-CCHC; 7.
 CC PRINTS; PR00939; C2HCZNFINGER.
 CC SMART; SM00343; Znf_C2HC; 7.
 CC PROSITE; PS00158; ZF_CCHC; 7.
 CC Zinc-finger; DNA-binding; Transcription regulation; Repressor; Repeat.
 KW ZN_FING 4 21 CCHC-TYPE 1.
 FT ZN_FING 45 62 CCHC-TYPE 2.
 FT ZN_FING 65 82 CCHC-TYPE 3.
 FT ZN_FING 89 106 CCHC-TYPE 4.
 FT ZN_FING 110 127 CCHC-TYPE 5.
 FT ZN_FING 128 145 CCHC-TYPE 6.
 FT ZN_FING 149 166 CCHC-TYPE 7.
 FT VARIANT 2 2 S -> R (IN CLONE 6).
 FT VARIANT 33 33 G -> GGGGTS (IN CLONE 4).
 FT VARIANT 65 65 D -> DE (IN CLONE 14).

FT VARIANT 98 98 G -> D (IN CLONE 14).
SQ SEQUENCE 170 AA; 18742 MW; 152BEC42881358E8 CRC64;

Query Match 75.0%; Score 21; DB 1; Length 170;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 QKXCXXX 8
|||
DB 122 QKCTKVK 129

RESULT 8

CNBP_CHKCK STANDARD; PRT; 172 AA.
AC O42395;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellular nucleic acid binding protein (CNBP).
GN ZNF9 OR CNBP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98129058; PubMed=9467877;
RA van Heumen W.R.A.; Claxton C.; Pickles J.O.;
RT "Sequence and tissue distribution of chicken cellular nucleic acid binding protein cDNA";
RL Comp. Biochem. Physiol. 118B:659-665(1997).
CC -!- FUNCTION: SINGLE STRANDED DNA-BINDING PROTEIN, WITH SPECIFICITY TO THE STEROL REGULATORY ELEMENT (SRE). CNBP IS INVOLVED IN STEROL-MEDIATED REPRESSION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ALSO PRESENT IN ENDOPLASMIC RETICULUM (BY SIMILARITY).
CC -!- SIMILARITY: TO S.POMBE BYR3 AND TO RETROVIRAL NUCLEIC ACID BINDING PROTEINS (NBP).
CC -!- SIMILARITY: Contains 7 CCHC-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; AF004942; AAB62243.1; -
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 7.
DR PRINTS; PM00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 7.
DR PROSITE; PSS0158; Znf_CCHC; 7.
KW Zinc-finger; DNA-binding; Transcription regulation; Repressor; Repeat.
FT ZN_FING 4 21 CCHC-TYPE 1.
FT ZN_FING 46 63 CCHC-TYPE 2.
FT ZN_FING 67 84 CCHC-TYPE 3.
FT ZN_FING 91 108 CCHC-TYPE 4.
FT ZN_FING 112 129 CCHC-TYPE 5.
FT ZN_FING 130 147 CCHC-TYPE 6.
FT ZN_FING 151 168 CCHC-TYPE 7.
SQ SEQUENCE 172 AA; 19043 MW; 123F4E248A980390 CRC64;

Query Match 75.0%; Score 21; DB 1; Length 172;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 QKXCXXX 8
|||
DB 124 QKCTKVK 131

RESULT 9

CNBP_HUMAN STANDARD; PRT; 177 AA.
AC P20694;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellular nucleic acid binding protein (CNBP).
GN ZNF9 OR CNBP.
OS Homo sapiens (Human), and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606, 10116;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-Human;
RX MEDLINE=89346750; PubMed=2562787;
RA Rajavashisth T.B.; Taylor A.K.; Andalibi A.; Svenson K.L.; Lusis A.J.;
RT "Identification of a zinc finger protein that binds to the sterol regulatory element";
RL Science 245:640-643(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-Human; TISSUE=Placenta;
RX MEDLINE=96011648; PubMed=7590281;
RA Flink I.L.; Workin E.;
RT "Organization of the gene encoding cellular nucleic acid-binding protein";
RL Gene 163:279-282(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES-Rat; STRAIN=Sprague-Dawley;
RX MEDLINE=95308326; PubMed=7788528;
RA Yasuda J.; Mashiya S.; Makino R.; Ohya S.; Sekiya T.; Hayashi K.;
RT "Cloning and characterization of rat cellular nucleic acid binding protein (CNBP) cDNA";
RL DNA Res. 2:45-49(1995).
CC -!- FUNCTION: SINGLE STRANDED DNA-BINDING PROTEIN, WITH SPECIFICITY TO THE STEROL REGULATORY ELEMENT (SRE). CNBP IS INVOLVED IN STEROL-MEDIATED REPRESSION.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ALSO PRESENT IN ENDOPLASMIC RETICULUM (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: PRESENT IN ALL TISSUES EXAMINED.
CC -!- SIMILARITY: TO S.POMBE BYR3 AND TO RETROVIRAL NUCLEIC ACID BINDING PROTEINS (NBP).
CC -!- SIMILARITY: Contains 7 CCHC-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; M28372; AAA61975.1; -
DR EMBL; U19765; AAA91782.1; -
DR EMBL; D45254; BAA08212.1; -
DR PIR; A32760; A32760.
DR PIR; JC2512; JC2512.
DR TRANSFAC; T02854; -
DR Genew; HGNC:13164; ZNF9.
DR MIM; 116955; -
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0006695; P:cholesterol biosynthesis; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 7.
DR PRINTS; PM00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 7.

DR PROSITE; PS0158; ZF_CCHC; 7.
 KW Zinc-finger; DNA-binding; Transcription regulation; Repressor; Repeat.
 FT ZN_FING 4 21 CCHC-TYPE 1.
 FT ZN_FING 52 69 CCHC-TYPE 2.
 FT ZN_FING 72 89 CCHC-TYPE 3.
 FT ZN_FING 96 113 CCHC-TYPE 4.
 FT ZN_FING 117 134 CCHC-TYPE 5.
 FT ZN_FING 135 152 CCHC-TYPE 6.
 FT ZN_FING 156 173 CCHC-TYPE 7.
 SQ SEQUENCE 177 AA; 19463 MW; 996F398285F52618 CRC64;

Query Match 75.0%; Score 21; DB 1; Length 177;
 Best Local Similarity 50.0%; Pred. No. 1.7e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
 ||| |
 Db 129 QKCTKVK 136

RESULT 10
 Y3_SOCMV STANDARD; PRT; 192 AA.
 AC P15633;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein 3 (ORF III).
 OS Soybean chlorotic mottle virus.
 OC Viruses; Retroid viruses; Caulimoviridae;
 OC Soybean chlorotic mottle-like viruses.
 OX NCBI_TaxID=10651;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90098857; PubMed=2602148;
 RA Hasegawa A., Verver J., Shimada A., Saito M., Goldbach R.,
 RA van Kammen A., Miki K., Kameya-Iwaki M., Hibi T.;
 RT "The complete sequence of soybean chlorotic mottle virus DNA and the
 RT identification of a novel promoter."
 RL Nucleic Acids Res. 17:9993-10013(1989).
 RN [2]
 RP REVISTONS TO 140 AND 164-167.
 RA Hibi T.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X15828; CAC16943.1;
 DR PIR; JS0373; JS0373.
 KW Hypothetical protein.
 SQ SEQUENCE 192 AA; 22484 MW; 8FDAC46B5D251096 CRC64;

Query Match 75.0%; Score 21; DB 1; Length 192;
 Best Local Similarity 50.0%; Pred. No. 1.8e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
 ||| |
 Db 153 QKACLDKFK 160

RESULT 11
 THIE_CAMJE STANDARD; PRT; 210 AA.
 AC Q9PNL3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Thiamine-phosphate pyrophosphorylase (EC 2.5.1.3) (TMP
 DE pyrophosphorylase) (TMP-PPase) (Thiamine-phosphate synthase).
 GN THE OR CUI081C.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Campylobacteraceae; Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences."
 RL Nature 403:665-668(2000).
 CC -!- FUNCTION: Condenses 4-methyl-5-(beta-hydroxyethyl)-thiazole
 CC monophosphate (THZ-P) and 4-amino-5-hydroxymethyl pyrimidine
 CC pyrophosphate (HMP-PP) to form thiamine monophosphate (TMP) (By
 CC similarity).
 CC -!- CATALYTIC ACTIVITY: 2-methyl-4-amino-5-hydroxymethylpyrimidine
 CC diphosphate + 4-4-methyl-5-(2-phosphonooxyethyl)-thiazole =
 CC diphosphate + thiamine monophosphate.
 CC -!- COFACTOR: Binds 1 magnesium ion per subunit (By similarity).
 CC -!- PATHWAY: Thiamine biosynthesis.
 CC -!- SIMILARITY: BELONGS TO THE TMP-PPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL139077; CAB73336.1;
 DR PIR; D81311; D81311.
 DR HSSP; P39594; 2TPS.
 DR HAMAP; MF_00097; -; 1.
 DR InterPro; IPR003733; TMP_synthase.
 DR Pfam; PF02581; TMP-TENI; 1.
 DR ProDom; PD149806; TMP_synthase; 1.
 DR TrRfam; TrR00693; thiE; 1.
 KW Thiamine biosynthesis; Transferase; Magnesium; Metal-binding;
 KW Complete proteome.
 FT METAL 72 91 MAGNESIUM (BY SIMILARITY).
 FT METAL 91 91 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 210 AA; 23001 MW; 753C83D43B0EB44A CRC64;

Query Match 75.0%; Score 21; DB 1; Length 210;
 Best Local Similarity 50.0%; Pred. No. 2e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
 ||| |
 Db 58 QKLCYSYK 65

RESULT 12
 PYRH_XYLFA STANDARD; PRT; 247 AA.
 ID PYRH_XYLFA
 AC Q9PEH0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Uridylate kinase (EC 2.7.4.-) (UK) (Uridine monophosphate kinase) (UMP
 DE kinase).
 GN PYRH OR XF1058.
 OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9aSC;
 RX MEDLINE=20365717; PubMed=10910347;
 RA SIMPSON A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Fracalanci A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fagundes J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohlseil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 CC -!- FUNCTION: Catalyzes the phosphorylation of UMP to UDP (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + UMP = ADP + UDP.
 CC -!- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; first step.
 CC -!- SIMILARITY: BELONGS TO THE UMP KINASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; AF003942; AAF83868.1;
 DR PIR; B82730; B82730.
 DR HAMAP; MF_01220; --:1.
 DR InterPro; IPR001048; Aa_kinase.
 DR Pfam; PF00696; Aa_kinase; 1.
 KW Transferase; Kinase; Pyrimidine biosynthesis; Complete proteome.
 SQ SEQUENCE 247 AA; 26827 MW; 4BD916D7692E0F9 CRC64;
 Query Match 75.0%; Score 21; DB 1; Length 247;
 Best Local Similarity 50.0%; Pred. No. 2.2e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 QKXCXXXK 8
 DB 177 QKHCDAVK 184
 RESULT 13
 FLIH_TREPA STANDARD; PRT; 309 AA.
 AC O83416;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Flagellar assembly protein flhH.
 GN FLIH OR TP0401.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.N., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis spirochete";
 RT Science 281:375-388(1998).
 RL SCIENCE 281:375-388(1998).
 CC -!- FUNCTION: NEEDED FOR FLAGELLAR REGROWTH AND ASSEMBLY (BY SIMILARITY).
 CC -!- SIMILARITY: TO FLIH IN OTHER BACTERIA AND ALSO TO FLII.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; AE001218; AAC65389.1;
 DR PIR; H71328; H71328.
 DR TIGR; TP0401;
 DR InterPro; IPR000563; Flag_FliH.
 DR Pfam; PF02108; FlhH; 1.
 KW Flagella; Complete proteome.
 SQ SEQUENCE 309 AA; 35339 MW; 213978124489C81F CRC64;
 Query Match 75.0%; Score 21; DB 1; Length 309;
 Best Local Similarity 50.0%; Pred. No. 2.7e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 QKXCXXXK 8
 DB 131 QKICDHSK 138
 RESULT 14
 DP3 YEAST STANDARD; PRT; 350 AA.
 ID DP3 YEAST
 AC P47110;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA polymerase delta subunit 3.
 GN POL32 OR YJR043C OR J1626
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=95397595; PubMed=7668047;
 RA Huang M.-E., Chuat J.-C., Galibert F.;
 RT "Analysis of a 42.5 kb DNA sequence of chromosome X reveals three tRNA genes and 14 new open reading frames including a gene most probably belonging to the family of ubiquitin-protein ligases";
 RT Yeast 11:775-781(1995).
 RL Yeast 11:775-781(1995).
 RN [2]
 RP CHARACTERIZATION.

```

RX MEDLINE=98344072; PubMed=9677405;
RA Gerik K.J., Li X., Pautz A., Burgers P.M.;
RT "Characterization of the two small subunits of Saccharomyces
RL cerevisiae DNA polymerase delta.";
RL J. Biol. Chem. 273:19747-19755(1998).
CC -!- SUBUNIT: HETEROTRIMER WITH SUBUNITS OF 125 KDA, 58 KDA AND 55 KDA.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L36344; AAA88745.1; -
DR EMBL; Z49543; CAAB9571.1; -
DR PIR; S57062; S57062.
DR SGD; S0003804; POL32.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0006298; P:mismatch repair; NAS.
KW DNA-directed DNA polymerase; DNA replication; Nuclear protein.
SQ SEQUENCE 350 AA; 40309 MW; DOB9CC52F26E20B2 CRC64;

Query Match 75.0%; Score 21; DB 1; Length 350;
Best Local Similarity 50.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXX 8
DB 103 QKCLTIK 110

RESULT 15
NK23_MOUSE
ID NK23_MOUSE STANDARD; PRT; 362 AA.
AC P97334; Q90260; Q9WV67;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein Nkx-2.3 (Homeobox protein Nkx-2 homolog C) (Nkx2-C)
DE (Homeobox protein Nkx-2.3 homolog 3).
GN NKX2-3 OR NKX-2.3 OR NKX2C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=Liver;
RX MEDLINE=97287401; PubMed=9142493;
RA Pabst O., Schneider A., Brand T., Arnold H.-H.;
RT "The mouse Nkx2-3 homeodomain gene is expressed in gut mesenchyme
RT during pre- and postnatal mouse development.";
RL Dev. Dyn. 209:29-35(1997).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=129;
RA Wang C.-C., Biben C., Robb R., Tarlinton D., Nassir F., Davidson N.O.,
RA Harvey R.P.;
RT "Homeodomain factor Nkx2-3 is required for normal development of the
RT gut-associated lymphoid tissue and the spleen.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=20387131; PubMed=10926756;
RA Wang C.-C., Biben C., Robb R., Nassir F., Barnett L., Davidson N.O.,
RA Koentgen F., Tarlinton D., Harvey R.P.;
RT "Homeodomain factor Nkx2-3 controls regional expression of leukocyte
RT homing coreceptor MadCAM-1 in specialized endothelial cells of the

```

```

RT Viscera.";
RL Dev. Biol. 224:152-167(2000).
RN [4]
RP FUNCTION.
RX MEDLINE=20253077; PubMed=10790368;
RA Pabst O., Foerster R., Lipp M., Engel H., Arnold H.-H.;
RT "Nkx2.3 is required for MadCAM-1 expression and homing of lymphocytes
RT in spleen and mucosa-associated lymphoid tissue.";
RL EMBO J. 19:2015-2023(2000).
RN [5]
RP TISSUE DISTRIBUTION.
RX MEDLINE=22136510; PubMed=12141427;
RA Biben C., Wang C.-C., Harvey R.P.;
RT "NK-2 class homeobox genes and pharyngeal/oral patterning: Nkx2-3 is
RT required for salivary gland and tooth morphogenesis.";
RL Int. J. Dev. Biol. 46:415-422(2002).
CC -!- FUNCTION: Transcriptional regulator essential for normal
CC development and functions of the small intestine and spleen.
CC Activates directly MadCAM1 expression. Required for homing of
CC lymphocytes in spleen and mucosa-associated lymphoid tissue. May
CC have a role during pharyngeal organogenesis.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: Expressed in spleen and intestine. Also
CC expressed in salivary gland and tongue, which are derivate of the
CC pharyngeal region.
CC -!- DEVELOPMENTAL STAGE: Expressed in gut mesenchyme during pre- and
CC postnatal development. Expressed as well in the pharyngeal floor
CC and pouches, and in the oral and branchial arch ectoderm.
CC Expression persisted in the developing thyroid until birth, in
CC mucous forming cells of salivary glands and in odontogenic
CC epithelium of the mandible.
CC -!- SIMILARITY: BELONGS TO THE NK-2 HOMEBOX FAMILY.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; Y11117; CAA72002.1; -
EMBL; AF202036; AAF08008.1; -
EMBL; AF155583; AAD38415.1; -
HSSP; P23441; 1FTT
MGP; MGI:97348; Nkx2-3.
InterPro; IPR001356; Homeobox.
InterPro; IPR000047; HTH_lambrepresr.
Pfam; PF00046; homeobox; 1.
PRINTS; PR00024; HOMEBOX.
PROSITE; PR00031; HTHREPRESSR.
PROSITE; PR00010; Homeobox; 1.
SMART; SM00389; HOX; 1.
PROSITE; PS00027; HOMEBOX_1; 1.
PROSITE; PS00071; HOMEBOX_2; 1.
Transcription regulation; Homeobox; DNA-binding;
Developmental protein; Nuclear protein.
DNA_BIND 145 204 HOMEBOX.
DOMAIN 59 67 POLY-GLU.
DOMAIN 216 220 POLY-PRO.
DOMAIN 268 283 POLY-ALA.
DOMAIN 295 302 POLY-GLY.
CONFLICT 44 44 C -> R (IN REF. 1).
CONFLICT 61 61 D -> E (IN REF. 1).
CONFLICT 132 132 K -> R (IN REF. 1).
CONFLICT 248 250 GVG -> RC (IN REF. 1).
CONFLICT 261 357 AYGVNGSAAAAAAYSGYCAYPGTGGGGG
GTASATTAQACSGFVNVLGFGSGGAQPLH
OGAAAGSACTGTLQ -> TATGTRPPQPPQOGLQ
ROLELRYVAVVAARPPRPPPCNPAPGDLRVTAG
IEAFVPLVR (IN REF. 1).
S -> A (IN REF. 3).

```

SQ SEQUENCE 362 AA: 38090 MW: 48886528EC3E1622 CRC64;

Query Match 75.0%; Score 21; DB 1; Length 362;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8

Db 116 QKSCQLKK 123

Search completed: July 29, 2003, 09:55:19
Job time : 10 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:52:37 ; Search time 32 Seconds
(without alignments)
64.513 Million cell updates/sec

Title: US-09-606-129A-17
Perfect score: 28
Sequence: 1 QKXCXXXK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phase:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_rvirus:.*
16: sp_bacteriap:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	82.1	543	10 Q9SH53	Q9sh53 arabidopsis
2	22	78.6	101	10 Q40843	Q40843 picea glauc
3	22	78.6	101	11 Q8CER5	Q8cer5 mus musculu
4	22	78.6	141	16 Q8F5B2	Q8f5b2 leptospira
5	22	78.6	149	16 Q9PMC2	Q9pmc2 campylobact
6	22	78.6	195	5 Q9NDT4	Q9ndt4 balanus amp
7	22	78.6	322	16 Q99X03	Q99x03 staphylococ
8	22	78.6	322	16 Q8NVK0	Q8nvk0 staphylococ
9	22	78.6	348	4 Q8WUF3	Q8wuf3 homo sapien
10	22	78.6	371	10 Q9LII9	Q9llli9 arabidopsis
11	22	78.6	419	10 Q9SGT3	Q9sgt3 arabidopsis
12	22	78.6	451	16 Q97FL8	Q97fl8 clostridium
13	22	78.6	458	5 Q95X02	Q95x02 naegleria f
14	22	78.6	484	5 Q9BKW1	Q9bkw1 naegleria f
15	22	78.6	541	4 Q8N2S8	Q8n2s8 homo sapien
16	22	78.6	564	12 Q9EN64	Q9en64 spodoptera

17	22	78.6	607	5 Q9V6S7	Q9v6s7 drosophila
18	22	78.6	617	5 Q16269	Q16269 caenorhabdi
19	22	78.6	642	5 Q76435	Q76435 caenorhabdi
20	22	78.6	683	12 Q36401	Q36401 alcellaphine
21	22	78.6	688	5 Q9N2S4	Q9n2s4 caenorhabdi
22	22	78.6	731	4 Q9NUH2	Q9nuh2 homo sapien
23	22	78.6	744	5 Q9VAG2	Q9vag2 drosophila
24	22	78.6	792	10 Q9AS80	Q9as80 oryza sativ
25	22	78.6	822	5 Q25021	Q25021 haemonchus
26	22	78.6	835	4 Q96J22	Q96j22 homo sapien
27	22	78.6	836	5 Q9NH46	Q9nh46 trypanosoma
28	22	78.6	840	10 Q93X57	Q93x57 fragaria an
29	22	78.6	845	5 Q9YLP8	Q9ylp8 plasmodium
30	22	78.6	870	5 Q76228	Q76228 trypanosoma
31	22	78.6	946	16 Q9ZK04	Q9zk04 helicobacte
32	22	78.6	2585	5 Q23587	Q23587 caenorhabdi
33	22	78.6	3488	2 Q8G986	Q8g986 planktothri
34	21	75.0	81	16 Q8FLV3	Q8flv3 leptospira
35	21	75.0	82	16 Q25977	Q25977 helicobacte
36	21	75.0	86	16 Q9ZJ10	Q9zj10 helicobacte
37	21	75.0	87	10 Q8VX81	Q8vx81 pinus pinas
38	21	75.0	87	16 Q8DTK4	Q8dtk4 streptococc
39	21	75.0	93	16 Q8EXG2	Q8exg2 leptospira
40	21	75.0	106	13 Q90698	Q90698 gallus galli
41	21	75.0	149	16 Q8ZLL5	Q8zll5 salmonella
42	21	75.0	166	4 Q8WW36	Q8ww36 homo sapien
43	21	75.0	167	4 Q96NV3	Q96nv3 homo sapien
44	21	75.0	168	13 Q91594	Q91594 xenopus lae
45	21	75.0	170	11 Q9D548	Q9d548 mus musculu

ALIGNMENTS

RESULT 1

Q9SH53 ID Q9SH53 PRELIMINARY: PRT: 543 AA.
AC Q9SH53;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE F22C12.21.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Khan S., Brooks S., Buehler E., Chao Q., Dunn P., Kim C.,
RA Walker M., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,
RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F22C12 from chromosome
I.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBDJ databases.
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL; AC007764; AAF24567.1, ..
DR HSSP; F01008; IATH.
DR InterPro; IPR000215; Serpin.
DR Pfam; PF00079; serpin; 2.
DR SMART; SM00093; SERPIN; 1.
DR PROSITE; PS00284; SERPIN; 1.
KW Protease inhibitor; Serine protease inhibitor; Serpin.
SQ SEQUENCE 543 AA; 61413 MW; 0FB606F0EB26FC7 CRC64;

Query Match 82.1%; Score 23; DB 10; Length 543;
Best Local Similarity 50.0%; Pred No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
|||

Db 34 QKLCSSK 41

RESULT 2

Q40843 ID Q40843 PRELIMINARY; PRT; 101 AA.
 AC Q40843;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Late embryogenesis abundant protein (Fragment).
 GN EMB16.
 OS Picea glauca (White spruce).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Picea.
 OX NCBI_TaxID=3330;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dong J.-Z., Dunstan D.I.;
 RT "Gene expression during somatic embryogenesis.";
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: L47629; AAB01553.1; --
 FT NON_TER 1
 SQ SEQUENCE 101 AA; 11554 MW; C41BC7180BF0566B CRC64;

Query Match 78.6%; Score 22; DB 10; Length 101;
 Best Local Similarity 50.0%; Pred. No. 1.7e+02;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
 || | |

Db 8 QKSCSDQK 15

RESULT 3

Q8CER5 ID Q8CER5 PRELIMINARY; PRT; 101 AA.
 AC Q8CER5;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=22354683; PubMed=12456851;
 RA The FANTOM Consortium.
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL: AK016134; BAC25474.1; --
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 101 AA; 10793 MW; CFD99672DCA89B3B CRC64;

Query Match 78.6%; Score 22; DB 11; Length 101;
 Best Local Similarity 50.0%; Pred. No. 1.7e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
 || | |

Db 14 QKDCLASK 21

RESULT 4

Q8F5B2 ID Q8F5B2 PRELIMINARY; PRT; 141 AA.
 AC Q8F5B2;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN LA1772.
 OS Leptospira interrogans.
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
 OX NCBI_TaxID=173;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
 RA Ren S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE011354; AAM48971.1; --
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 141 AA; 16495 MW; 0C84C41037C83035 CRC64;

Query Match 78.6%; Score 22; DB 16; Length 141;
 Best Local Similarity 50.0%; Pred. No. 2.2e+02;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
 || | |

Db 23 QKACNTLK 30

RESULT 5

Q9PMC2 ID Q9PMC2 PRELIMINARY; PRT; 149 AA.
 AC Q9PMC2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Homolog of BLC protein.
 GN CJ1547.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Campylobacteraceae; Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 111168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 DR EMBL: ALI39078; CAB73963.1; --
 DR InterPro; IPR000566; Lipocln_cytFABP.
 DR PROSITE; PS00213; LIPOCALIN; 1.
 KW Complete proteome.
 SQ SEQUENCE 149 AA; 17432 MW; 757F171A92AE9C10 CRC64;

Query Match 78.6%; Score 22; DB 16; Length 149;
 Best Local Similarity 50.0%; Pred. No. 2.3e+02;

Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
 || | |

Db 28 QKSCLSNR 35

RESULT 6

Q9NDT4 ID Q9NDT4 PRELIMINARY; PRT; 195 AA.
 AC Q9NDT4;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

```
DE BCS-4.
GN BCS-4.
OS Balanus amphitrite (Barnacle).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Maxillopoda; Cirripedia;
OC Thoracica; Sessilia; Balanomorpha; Balanoidea; Balanidae; Balanus.
OX NCBI_TaxID=32267;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20314484; PubMed=10854786;
RA Okazaki Y., Shizuri Y.;
RT "Structures of six cDNAs expressed specifically at cypris larvae of
RT barnacles, Balanus amphitrite.";
RL Gene 250:127-135(2000).
DR EMBL; AB021905; BAA99546.1; -.
DR InterPro; IPR003645; FOLN.
DR SMART; SM00274; FOLN; 2.
SQ SEQUENCE 195 AA; 20545 MW; 72204A9304191ED1 CRC64;

Query Match 78.6%; Score 22; DB 5; Length 195;
Best Local Similarity 50.0%; Pred. No. 2.8e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
DB 86 QKTCADIK 93

RESULT 7
Q99X03 PRELIMINARY; PRT; 322 AA.
AC Q99X03;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Hypothetical protein SAV0219.
GN SAV0219 OR SA0212.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hiraoka H., Kihara S., Goto S., Yabuuchi J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AF003358; BAB56381.1; -.
DR EMBL; AF003129; BAB41434.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 322 AA; 36655 MW; 6089E05E6D731600 CRC64;

Query Match 78.6%; Score 22; DB 16; Length 322;
Best Local Similarity 50.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
DB 301 QKACOTLK 308

RESULT 8
Q8NYK0 PRELIMINARY; PRT; 322 AA.
ID Q8NYK0;
AC Q8NYK0;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
```

```
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Hypothetical protein MW0195.
GN MW0195.
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
DR EMBL; AP004822; BAB94060.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 322 AA; 36671 MW; 1289EDA9C9A87E1A CRC64;

Query Match 78.6%; Score 22; DB 16; Length 322;
Best Local Similarity 50.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
DB 301 QKACOTLK 308

RESULT 9
Q8WUF3 PRELIMINARY; PRT; 348 AA.
ID Q8WUF3;
AC Q8WUF3;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020595; RAH20595.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 348 AA; 35999 MW; B946D95225DACALB CRC64;

Query Match 78.6%; Score 22; DB 4; Length 348;
Best Local Similarity 50.0%; Pred. No. 4.4e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
DB 260 QKDCLASK 267

RESULT 10
Q9LI19 PRELIMINARY; PRT; 371 AA.
ID Q9LI19;
AC Q9LI19;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Genomic DNA, chromosome 3, TAC clone:k24A2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Columbia;
RA  Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Columbia;
RX  MEDLINE=20363099; PubMed=10907853;
RA  Nakamura Y.;
RT  "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT  Sequence features of the regions of 4,251,695 bp covered by ninety pl,
RT  TAC and BAC clones."
RL  DNA Res. 7:217-221(2000).
DR  EMBL; AP001302; BAB01482.1; -
DR  InterPro; IPR001087; Lipase_GDSL.
DR  Pfam; PF00657; Lipase_GDSL; 1.
SQ  SEQUENCE 371 AA; 40778 MW; 0F67C0F4AEDFEADE CRC64;

Query Match 78.6%; Score 22; DB 10; Length 371;
Best Local Similarity 50.0%; Pred. No. 4.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
Db 363 QKACKLTK 370

RESULT 11
Q9SGT3 PRELIMINARY; PRT; 419 AA.
ID Q9SGT3
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE T6H22.14 protein.
GN T6H22.14
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eutroids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Nguyen M., Lam B., Buehler E., Dunn P., Gonzalez A.,
RA Kremenetskaia I., Kim C., Lenz C., Li J., Liu S., Luros S.,
RA Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S., Walker M., Yu G.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009894; AAF02843.1; -
DR InterPro; IPR004396; Cons_hypoth92.
DR PRINTS; PR00326; GTP1_OBG.
DR TIGRFAMS; TIGR00092; TIGR00092; 1.
SQ SEQUENCE 419 AA; 45401 MW; B3212C374F2A963B CRC64;

Query Match 78.6%; Score 22; DB 10; Length 419;
Best Local Similarity 50.0%; Pred. No. 5e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
Db 42 QKRCFSSK 49

RESULT 12
Q97FL8 PRELIMINARY; PRT; 451 AA.
ID Q97FL8
AC Q97FL8;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)

```

```

DE Ethanolamine ammonia lyase large subunit.
GN CAC2718.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007769; AAK80664.1; -
KW Lyase; Complete proteome.
SQ SEQUENCE 451 AA; 50181 MW; 70DD188A633D57F2 CRC64;

Query Match 78.6%; Score 22; DB 16; Length 451;
Best Local Similarity 50.0%; Pred. No. 5.3e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
Db 137 QKICNTAK 144

RESULT 13
Q95X02 PRELIMINARY; PRT; 458 AA.
ID Q95X02
AC Q95X02;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Naegleria fowleri.
GN NP-B.
OS Naegleria fowleri.
OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.
OX NCBI_TaxID=5763;
RN [1]
RP SEQUENCE FROM N.A.
RA Nickel R., Leippe M.;
RT "Pore-forming peptides of Naegleria fowleri."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF196309; AAL01158.1; -
DR InterPro; IPR000004; SapB.
DR ProDom; PD001732; SapB; 4.
DR SMART; SM00118; SAPB; 5.
FT NON_TER 1
FT SEQUENCE 458 AA; 50452 MW; 433F60CDD6E4C904 CRC64;

Query Match 78.6%; Score 22; DB 5; Length 458;
Best Local Similarity 50.0%; Pred. No. 5.4e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
Db 338 QKVCSTLK 345

RESULT 14
Q9BKMI PRELIMINARY; PRT; 484 AA.
ID Q9BKMI
AC Q9BKMI;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Naegleria fowleri.
GN PRONP-B.
OS Naegleria fowleri.

```

OC Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria.
OX NCBI_TaxID=5763;

RN [1]
RP SEQUENCE FROM N.A.
RA Nickel R., Benkert C., Jacobs T., Marti T., Marciano-Cabral F.,
RA Leippe M.:
RT "Pore-forming peptides of Naegleria fowleri.";
RL submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF154047; AAK21659.1; -
DR InterPro: IPR000004; SapB.
DR ProDom: PD001732; SapB; 4.
DR SMART: SM00118; SapB; 5.
SQ SEQUENCE 484 AA; 53237 MW; 63DA3AAE7CA578B0 CRC64;

Query Match 78.6%; Score 22; DB 5; Length 484;
Best Local Similarity 50.0%; Pred. No. 5.6e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
DB 364 QKVCSTLK 371

RESULT 15

Q8N2S8 PRELIMINARY; PRT; 541 AA.
AC Q8N2S8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ90005.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yanamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK074486; BAC11017.1; -
DR InterPro: IPR002867; Znf_C6HC.
DR Pfam: PF01485; IPR; 1.
DR SMART: SM00647; IPR; 2.
KW Hypothetical protein.
SQ SEQUENCE 541 AA; 57524 MW; E38D17A1447422BA CRC64;

Query Match 78.6%; Score 22; DB 4; Length 541;
Best Local Similarity 50.0%; Pred. No. 6.1e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
||| |
DB 453 QKCLASK 460

Search completed: July 29, 2003, 09:56:28
Job time : 35 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 29, 2003, 09:51:58 ; Search time 34.0667 Seconds
(without alignments)
32.615 Million cell updates/sec

Title: US-09-606-129A-16
Perfect score: 37
Sequence: 1 KRRIXHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	94.6	295	24	ABB99870 Rat biliverdin red
2	35	94.6	295	24	ABB99874 Rat biliverdin red
3	35	94.6	295	24	ABB99875 Rat biliverdin red
4	35	94.6	295	24	ABB99876 Rat biliverdin red
5	35	94.6	295	24	ABB99877 Rat biliverdin red
6	35	94.6	296	24	ABB99871 Human biliverdin r
7	34	91.9	40	17	AAR93552 Bovine FKBP-13 imm
8	34	91.9	83	22	AAG76114 Human colon cancer
9	34	91.9	99	13	AAR28979 Bovine RFBKP. Bos

10	34	91.9	104	21	AAG03758 Human secreted pro
11	34	91.9	137	22	AAB87656 Bovine mammary tis
12	34	91.9	141	17	AAR93551 Human FKBP-13 immu
13	34	91.9	142	24	ABP75992 Human GENSET prote
14	34	91.9	142	24	ABP76156 Human GENSET prote
15	34	91.9	163	23	ABG60082 Human DITHP polype
16	34	91.9	182	22	AAU31050 Novel human secret
17	34	91.9	795	22	AAU29777 Novel human secret
18	33	89.2	285	21	AAG28710 Arabidopsis thalia
19	33	89.2	473	24	ABB82773 Mouse OAS L protei
20	33	89.2	510	23	ABB90889 Herbicidally activ
21	32	86.5	109	22	ABB57919 Drosophila melanog
22	31	83.8	94	21	AAU97243 Murine INGB1 isofo
23	31	83.8	279	21	AAU97242 Murine P37ING1 pol
24	31	83.8	853	22	ABB64007 Drosophila melanog
25	30	81.1	46	22	ABP35789 PSK-H2 positioins p
26	30	81.1	83	22	ABG14241 Novel human diagno
27	30	81.1	103	22	ABG08440 Novel human diagno
28	30	81.1	141	21	AAB53247 Human colon cancer
29	30	81.1	197	24	ABU02223 S. pneumoniae type
30	30	81.1	352	22	ABB62968 Drosophila melanog
31	30	81.1	504	18	AAW36140 Bovine P58 protein
32	30	81.1	504	22	ABG09790 Novel human diagno
33	30	81.1	508	21	AAG48146 Arabidopsis thalia
34	30	81.1	508	23	ABB91938 Arabidopsis thalia
35	30	81.1	526	23	ABG30696 Murine mBYRK2 poly
36	30	81.1	527	23	ABB80743 Rat YAK1 protein k
37	30	81.1	528	19	AAW64559 Human protein kina
38	30	81.1	528	22	AAE02011 Human YAK1 (YAK1)
39	30	81.1	533	21	AAG48145 Arabidopsis thalia
40	30	81.1	534	22	ABB71968 Drosophila melanog
41	30	81.1	549	21	AAU68781 Amino acid sequenc
42	30	81.1	564	21	AAG48144 Arabidopsis thalia
43	30	81.1	565	20	AAW96316 Acidic leucine ami
44	30	81.1	571	20	AAW96315 Acidic leucine ami
45	29	78.4	44	22	AAU20305 Human novel endocr

ALIGNMENTS

RESULT 1
ABB99870
ID ABB99870 standard; Protein; 295 AA.
XX
AC ABB99870;
XX
DT 24-JAN-2003 (first entry)
XX
DE Rat biliverdin reductase (BVR).
XX
KW Rat; biliverdin reductase; BVR; bilirubin; protein co-ordinate data;
KW three dimensional structure; 3D structure; X-ray crystallography;
KW virtual drug screening; rational drug design; drug discovery;
KW recombinant expression; Escherichia coli; EC 1.3.1.24; enzyme.
XX
OS Rattus norvegicus.
XX
FH Key Location/Qualifiers
FT Region 9..14 /label= Beta_sheet_1
FT Region 15..20 /note= "Dinucleotide binding protein fingerprint"
FT Region 18..27 /label= Alpha_helix_A
FT Region 30..32 /label= Alpha_helix_B
FT Region 36..42 /label= Beta_sheet_2
FT Binding-site 44..46 /label= Nucleotide_binding_pocket
FT /note= "Peptides containing these residues are specifically claimed"

```
FT Binding-site 44..45 /note= "Binds NADPH"
FT Binding-site 46 /note= "Binds NADH"
FT Region 49..50 /label= Beta_sheet_3
FT Region 52..54 /label= Beta_sheet_4
FT Region 57..63 /label= Alpha_helix_C
FT Region 69..72 /label= Beta_sheet_5
FT Binding-site 73 /note= "Substrate/cofactor binding site"
FT Region 76..88 /label= Alpha_helix_D
FT Region 92..96 /label= Beta_sheet_6
FT Binding-site 96 /note= "Binds NADPH"
FT Active-site 96..98 /label= Active_site_EYP_motif
FT /note= "Peptides containing these residues are specifically claimed"
FT Region 103..116 /label= Alpha_helix_E
FT Region 119..122 /label= Beta_sheet_7
FT Region 126..128 /label= 3-10_helix_F
FT Region 130..139 /label= Alpha_helix_G
FT Region 144..153 /label= Beta_sheet_8
FT Region 157..160 /label= Alpha_helix_H
FT Region 164..167 /label= Alpha_helix_I
FT Region 169..179 /label= Alpha_helix_J
FT Binding-site 171 /note= "Binds NADPH"
FT /note= "Peptides containing this residue are specifically claimed"
FT Region 183..192 /label= Beta_sheet_9
FT Region 193..195 /label= 3-10_helix_K
FT Region 197..204 /label= Beta_sheet_10
FT Region 211..218 /label= Beta_sheet_11
FT Binding-site 218 /note= "Binds NADPH"
FT /note= "Peptides containing this residue are specifically claimed"
FT Region 224..231 /label= Beta_sheet_12
FT Binding-site 224 /note= "Binds NADPH"
FT Binding-site 226 /note= "Binds NADPH"
FT /note= "Peptides containing this residue are specifically claimed"
FT Region 235..237 /label= Beta_sheet_13
FT Region 250..261 /label= Alpha_helix_L
FT Region 266..291 /label= Alpha_helix_M
FT Region 279..292 /note= "Zinc co-ordinating motif"
XX

PN JP2002238553-A.
XX
PD 27-AUG-2002.
XX
XX 19-FEB-2001; 2001JP-0042501.
XX
XX 19-FEB-2001; 2001JP-0042501.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
XX
XX WPI; 2003-003949/01.
XX N-PSDB; ABQ79296.
XX
XX Steric structure of biliverdin reductase (BVR) used for screening of
XX drugs, drug design and peptide fragment defined by particular amino
XX acid sequences
XX
XX Claim 1; Fig 4; 72pp; Japanese.
XX
XX The invention relates to the elucidation of the three dimensional
XX structure of rat biliverdin reductase (ABB99870). Rat BVR is structurally
XX similar to human BVR (ABB99871) and Synecocystis sp. PCC 6803 BVR
XX (ABB99872). BVR (EC 1.3.1.24) catalyses the conversion of biliverdin to
XX bilirubin, using NADH or NADPH as an electron donor. The BVR crystal
XX structure, and peptide fragments of rat BVR containing residues 44-46,
XX 96-98, 171, 218, 224 or 226 may be used for the virtual screening and
XX design of drugs which target BVR. The present sequence represents rat
XX BVR, which was recombinantly expressed in Escherichia coli BL21,
XX crystallised in 0.2M sodium potassium tartrate with 2.0M of ammonium
XX sulphate as a precipitating agent, and subjected to X-ray
XX crystallography.
XX
XX Sequence 295 AA;
XX
XX Query Match 94.6%; Score 35; DB 24; Length 295;
XX Best Local Similarity 85.7%; Pred. No. 27;
XX Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 KKRIKHC 7
XX Db 274 KKRIMHC 280
XX
XX RESULT 2
XX ABB99874
XX ID ABB99874 standard; Protein; 295 AA.
XX
XX AC ABB99874;
XX
XX XX 24-JAN-2003 (first entry)
XX
XX DE Rat biliverdin reductase (BVR) mutant, E96A.
XX
XX KW Rat; biliverdin reductase; BVR; bilirubin; protein co-ordinate data;
XX three dimensional structure; 3D structure; X-ray crystallography;
XX virtual drug screening; rational drug design; drug discovery;
XX recombinant expression; Escherichia coli; EC 1.3.1.24; enzyme;
XX mutant; mutain.
XX
XX OS Rattus norvegicus.
XX
XX XX Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Region 9..14 /label= Beta_sheet_1
XX FT Region 15..20 /note= "Dinucleotide binding protein fingerprint"
XX FT Region 18..27 /label= Alpha_helix_A
XX FT Region 30..32 /label= Alpha_helix_B
XX FT Region 36..42 /label= Beta_sheet_2
XX FT
```

FT Binding-site 44..46
/label= Nucleotide_binding_pocket
/note= "Peptides containing these residues are
specifically claimed"

FT Binding-site 44..45
/note= "Binds NADPH"

FT Binding-site 46
/note= "Binds NADH"

FT Region 49..50
/label= Beta_sheet_3

FT Region 52..54
/label= Beta_sheet_4

FT Region 57..63
/label= Alpha_helix_C

FT Region 69..72
/label= Beta_sheet_5

FT Binding-site 73
/note= "Substrate/cofactor binding site"

FT Region 76..88
/label= Alpha_helix_D

FT Region 92..96
/label= Beta_sheet_6

FT Misc-difference 96
/note= "Ala replaces wild-type NADPH-binding Glu"

FT Active-site 96..98
/label= Active_site_EYP_motif
/note= "Peptides containing the wild-type residues are
specifically claimed"

FT Region 103..116
/label= Alpha_helix_E

FT Region 119..122
/label= Beta_sheet_7

FT Region 126..128
/label= 3-10_helix_F

FT Region 130..139
/label= Alpha_helix_G

FT Region 144..153
/label= Beta_sheet_8

FT Region 157..160
/label= Alpha_helix_H

FT Region 164..167
/label= Alpha_helix_I

FT Region 169..179
/label= Alpha_helix_J

FT Binding-site 171
/note= "Binds NADPH"

FT Region 183..192
/label= Beta_sheet_9

FT Region 193..195
/label= 3-10_helix_K

FT Region 197..204
/label= Beta_sheet_10

FT Region 211..218
/label= Beta_sheet_11

FT Binding-site 218
/note= "Binds NADPH"

FT Region 224..231
/label= Beta_sheet_12

FT Binding-site 224
/note= "Binds NADPH"

FT Binding-site 226
/note= "Binds NADPH"

FT Region 235..237
/label= Beta_sheet_13

FT Region 250..261
/label= Alpha_helix_L

FT Region 266..291

FT Region
/label= Alpha_helix_M
279..292
/note= "Zinc co-ordinating motif"

XX JP2002238553-A.
PN
XX
PD 27-AUG-2002.
XX
XX 19-FEB-2001; 2001JP-0042501.
XX PF
XX PR 19-FEB-2001; 2001JP-0042501.
XX PA (RIKA) RIKAGAKU KENKYUSHO.
XX DR WPI; 2003-003949/01.
XX
PT Steric structure of biliverdin reductase (BVR) used for screening of
PT drugs, drug design and peptide fragment defined by particular amino
PT acid sequences
XX
XX PS
XX PS Example 1; Page -; 72pp; Japanese.
XX
CC The invention relates to the elucidation of the three dimensional
CC structure of rat biliverdin reductase (ABB99870). Rat BVR is structurally
CC similar to human BVR (ABB99871) and Synecocystis sp. PCC 6803 BVR
CC (ABB99872). BVR (EC 1.3.1.24) catalyses the conversion of biliverdin to
CC bilirubin, using NADH or NADPH as an electron donor. The BVR crystal
CC structure, and peptide fragments of rat BVR containing residues 44-46,
CC 96-98, 171, 218, 224 or 226 may be used for the virtual screening and
CC design of drugs which target BVR. Sequences ABO99873-ABO99878
CC represent mutants of rat BVR referred to in the exemplification of the
CC invention.
CC Note: The present sequence is not given in the specification, but has
CC been derived from the sequence shown in Figure 4 and the information
CC given on page 64.
XX
SQ Sequence 295 AA;

Query Match 94.6%; Score 35; DB 24; Length 295;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative. 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIKHC 7
||| |
Db 274 KKRIMHC 280

RESULT 3
ABB99875
ID ABB99875 standard; Protein; 295 AA.
XX
XX AC ABB99875;
XX
DT 24-JAN-2003 (first entry)
XX
DE Rat biliverdin reductase (BVR) mutant, C73A.
XX
KW Rat; biliverdin reductase; BVR; bilirubin; protein co-ordinate data;
KW three dimensional structure; 3D structure; X-ray crystallography;
KW virtual drug screening; rational drug design; drug discovery;
KW recombinant expression; Escherichia coli; EC 1.3.1.24; enzyme;
KW mutant; mutein.
XX
OS Rattus norvegicus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Region 9..14
FT /label= Beta_sheet_1
FT Region 15..20
FT /note= "Dinucleotide binding protein fingerprint"
FT Region 18..27
FT /label= Alpha_helix_A

FT Region 30..32 /label= Alpha_helix_B
FT Region 36..42 /label= Beta_sheet_2
FT Binding-site 44..46 /label= Nucleotide_binding_pocket
FT /note= "Peptides containing these residues are
FT specifically claimed"
FT Binding-site 44..45 /note= "Binds NADPH"
FT Binding-site 46 /note= "Binds NADH"
FT Region 49..50 /label= Beta_sheet_3
FT Region 52..54 /label= Beta_sheet_4
FT Region 57..63 /label= Alpha_helix_C
FT Region 69..72 /label= Beta_sheet_5
FT Misc-difference 73 /note= "Ala replaces wild-type substrate/cofactor-
FT binding Cys"
FT Region 76..88 /label= Alpha_helix_D
FT Region 92..96 /label= Beta_sheet_6
FT Binding-site 96 /note= "Binds NADPH"
FT Active-site 96..98 /label= "Active_site_EYP_motif
FT /note= "Peptides containing these residues are
FT specifically claimed"
FT Region 103..116 /label= Alpha_helix_E
FT Region 119..122 /label= Beta_sheet_7
FT Region 126..128 /label= 3-10_helix_F
FT Region 130..139 /label= Alpha_helix_G
FT Region 144..153 /label= Beta_sheet_8
FT Region 157..160 /label= Alpha_helix_H
FT Region 164..167 /label= Alpha_helix_I
FT Region 169..179 /label= Alpha_helix_J
FT Binding-site 171 /note= "Binds NADPH"
FT /note= "Peptides containing this residue are
FT specifically claimed"
FT Region 183..192 /label= Beta_sheet_9
FT Region 193..195 /label= 3-10_helix_K
FT Region 197..204 /label= Beta_sheet_10
FT Region 211..218 /label= Beta_sheet_11
FT Binding-site 218 /note= "Binds NADPH"
FT /note= "Peptides containing this residue are
FT specifically claimed"
FT Region 224..231 /label= Beta_sheet_12
FT Binding-site 224 /note= "Binds NADPH"
FT Binding-site 226 /note= "Binds NADPH"
FT /note= "Binds NADPH"
FT /note= "Peptides containing this residue are
FT specifically claimed"

FT Region 235..237 /label= Beta_sheet_13
FT Region 250..261 /label= Alpha_helix_L
FT Region 266..291 /label= Alpha_helix_M
FT Region 279..292 /note= "Zinc co-ordinating motif"
XX JP2002238553-A.
XX 27-AUG-2002.
XX 19-FEB-2001; 2001JP-0042501.
XX 19-FEB-2001; 2001JP-0042501.
XX (RIKA) RIKAGAKU KENKYUSHO.
XX WPI; 2003-003949/01.
XX Steric structure of biliverdin reductase (BVR) used for screening of
FT drugs, drug design and peptide fragment defined by particular amino
FT acid sequences
XX Example 1; Page 72pp; Japanese.
XX The invention relates to the elucidation of the three dimensional
CC structure of rat biliverdin reductase (AB99870). Rat BVR is structurally
CC similar to human BVR (AB99871) and Synchocystis sp. PCC 6803 BVR
CC (AB99872). BVR (EC 1.3.1.24) catalyses the conversion of biliverdin to
CC bilirubin, using NADH or NADPH as an electron donor. The BVR crystal
CC structure, and peptide fragments of rat BVR containing residues 44-46,
CC 96-98, 171, 218, 224 or 226 may be used for the virtual screening and
CC design of drugs which target BVR. Sequences AB99873-AB99878
CC represent mutants of rat BVR referred to in the exemplification of the
CC the present sequence represents a rat BVR mutant in which the Cys
CC at position 73 is replaced by Ala. Synchocystis sp. PCC 6803 BVR
CC (AB99872) also has an Ala at the equivalent position.
CC Note: The present sequence is not given in the specification, but has
CC been derived from the sequence shown in Figure 4 and the information
CC given on page 63.
XX SQ Sequence 295 AA;
Query Match 94.6%; Score 35; DB 24; Length 295;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 KKRHXHC 7
Db 274 KKRIMHC 280
RESULT 4
AB99876
ID AB99876 standard; Protein; 295 AA.
XX AC AB99876;
XX 24-JAN-2003 (first entry)
XX Rat biliverdin reductase (BVR) mutant, E46A.
XX Rat; biliverdin reductase; BVR; bilirubin; protein co-ordinate data;
KW three dimensional structure; 3D structure; X-ray crystallography;
KW virtual drug screening; rational drug design; drug discovery;
KW recombinant expression; Escherichia coli; EC 1.3.1.24; enzyme;
XX mutant; mutein.
XX Rattus norvegicus.
OS Synthetic.
XX

FT	Binding-site	224	/note= "Binds NADPH"	
FT	Binding-site	226	/note= "Binds NADPH"	
FT		/note= "Peptides containing this residue are		
FT		/note= specifically claimed"		
FT	Region	235..237		
FT		/label= Beta_sheet_l3		
FT	Region	250..261		
FT		/label= Alpha_helix_L		
FT	Region	266..291		
FT		/label= Alpha_helix_M		
FT	Region	279..292		
FT		/note= "zinc co-ordinating motif"		
XX				
PN	JP2002238553-A.			
XX				
PD	27-AUG-2002.			
XX				
XX	19-FEB-2001; 2001JP-0042501.			
XX				
PR	19-FEB-2001; 2001JP-0042501.			
XX				
PA	(RIKA) RIKAGAKU KENKYUSHO.			
XX				
DR	WPI; 2003-003949/01.			
XX				
CC	Steric structure of biliverdin reductase (BVR) used for screening of			
PT	drugs, drug design and peptide fragment defined by particular amino			
PT	acid sequences -			
XX				
PS	Example 1; Page -; 72pp; Japanese.			
XX				
CC	The invention relates to the elucidation of the three dimensional			
CC	structure of rat biliverdin reductase (ABB99870). Rat BVR is structurally			
CC	similar to human BVR (ABB99871) and Synchocystis sp. PCC 6803 BVR			
CC	(ABB99872). BVR (EC 1.3.1.24) catalyses the conversion of biliverdin to			
CC	bilirubin, using NADH or NADPH as an electron donor. The BVR crystal			
CC	structure, and peptide fragments of rat BVR containing residues 44-46,			
CC	96-98, 171, 218, 224 or 226 may be used for the virtual screening and			
CC	design of drugs which target BVR. Sequences ABQ99873-ABQ99878			
CC	represent mutants of rat BVR referred to in the exemplification of the			
CC	invention.			
CC	Note: The present sequence is not given in the specification, but has			
CC	been derived from the sequence shown in Figure 4 and the information			
CC	given on page 65.			
XX				
SQ	Sequence 295 AA;			
	Query Match	94.6%;	Score 35;	DB 24; Length 295;
	Best Local Similarity	85.7%;	Pred. No. 27;	
	Matches 6; Conservative	0;	Mismatches 1;	Indels 0; Gaps 0
Qy	1 KKRXXHC 7			
Db	274 KKRIMHC 280			
RESULT 5				
ABB99877				
ID	ABB99877 standard; Protein; 295 AA.			
XX				
AC	ABB99877;			
XX				
DT	24-JAN-2003 (first entry)			
XX				
DE	Rat biliverdin reductase (BVR) mutant, R44A/R45A.			
XX				
KW	Rat; biliverdin reductase; BVR; bilirubin; protein co-ordinate data;			
KW	three dimensional structure; 3D structure; X-ray crystallography;			
KW	virtual drug screening; rational drug design; drug discovery;			
KW	recombinant expression; Escherichia coli; EC 1.3.1.24; enzyme;			
KW	mutant; muten.			

XX Human; biliverdin reductase; BVR; bilirubin; protein co-ordinate data;
KW three dimensional structure; 3D structure; X-ray crystallography;
KW virtual drug screening; rational drug design; drug discovery;
KW EC 1.3.1.24; enzyme.

XX Homo sapiens.

XX Key Location/Qualifiers
FH 9..14
FT Region /label= Beta_sheet_1
FT Region 15..20
FT Region /note= "dinucleotide binding protein fingerprint"
FT Region 18..27
FT Region /label= Alpha_helix_A
FT Region 30..33
FT Region /label= Alpha_helix_B
FT Region 37..43
FT Region /label= Beta_sheet_2
FT Binding-site 45..47
FT Binding-site /label= Nucleotide_binding_pocket
FT Binding-site 45..46
FT Binding-site /note= "Binds NADPH"
FT Binding-site 47
FT Binding-site /note= "Binds NADH"
FT Region 50..51
FT Region /label= Beta_sheet_3
FT Region 53..55
FT Region /label= Beta_sheet_4
FT Region 58..64
FT Region /label= Alpha_helix_C
FT Region 70..73
FT Region /label= Beta_sheet_5
FT Binding-site 74
FT Binding-site /note= "Substrate/cofactor binding site"
FT Region 77..89
FT Region /label= Alpha_helix_D
FT Region 93..97
FT Binding-site /label= Beta_sheet_6
FT Binding-site 97
FT Binding-site /note= "Binds NADPH"
FT Active-site 97..99
FT Region /label= Active_site_EYP_motif
FT Region 104..117
FT Region /label= Alpha_helix_E
FT Region 120..123
FT Region /label= Beta_sheet_7
FT Region 127..129
FT Region /label= 3-10_helix_F
FT Region 131..140
FT Region /label= Alpha_helix_G
FT Region 145..154
FT Region /label= Beta_sheet_8
FT Region 158..161
FT Region /label= Alpha_helix_H
FT Region 165..168
FT Region /label= Alpha_helix_I
FT Region 170..180
FT Region /label= Alpha_helix_J
FT Binding-site 172
FT Binding-site /note= "Binds NADPH"
FT Region 184..193
FT Region /label= Beta_sheet_9
FT Region 194..196
FT Region /label= 3-10_helix_K
FT Region 198..205
FT Region /label= Beta_sheet_10
FT Region 212..219
FT Region /label= Beta_sheet_11
FT Binding-site 219
FT Binding-site /note= "Binds NADPH"
FT Region 225..232
FT Region /label= Beta_sheet_12

FT Binding-site 225 /note= "Binds NADPH"
FT Binding-site 227 /note= "Binds NADPH"
FT Region 236..238
FT Region /label= Beta_sheet_13
FT Region 251..262
FT Region /label= Alpha_helix_L
FT Region 267..292
FT Region /label= Alpha_helix_M
FT Region 280..293
FT Region /note= "zinc co-ordinating motif"
XX JP2002238553-A.
XX 27-AUG-2002.
XX 19-FEB-2001; 2001JP-0042501.
XX 19-FEB-2001; 2001JP-0042501.
XX (RIKA) RIKAGAKU KENKYUSHO.
XX WPI; 2003-003949/01.
XX Steric structure of biliverdin reductase (BVR) used for screening of
FT drugs, drug design and peptide fragment defined by particular amino
FT acid sequences -
XX Example 1; Fig 4; 72pp; Japanese.
XX The invention relates to the elucidation of the three dimensional
CC structure of rat biliverdin reductase (ABB99870). Rat BVR is structurally
CC similar to human BVR (ABB99871) and Synchocystis sp. PCC 6803 BVR
CC (ABB99872). BVR (EC 1.3.1.24) catalyses the conversion of biliverdin to
CC bilirubin, using NADH or NADPH as an electron donor. The BVR crystal
CC structure, and peptide fragments of rat BVR containing residues 44-46,
CC 96-98, 171, 218, 224 or 226 may be used for the virtual screening and
CC design of drugs which target BVR. The present sequence represents human
CC BVR, which has a 3D structure similar to that of rat biliverdin
CC (ABB99870).
XX Sequence 296 AA;
SQ Query Match 94.6%; Score 35; DB 24; Length 296;
Best Local Similarity 85.7%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 KKRIXHC 7
Db 275 KKRLHC 281
IIII II
RESULT 7
AAR93552
ID AAR93552 standard; Peptide; 40 AA.
XX AAR93552;
AC AAR93552;
XX 25-JUN-1996 (first entry)
XX Bovine FKBP-13 immunophilin N-terminal peptide.
XX FKBP-13; immunophilin; FK506; rapamycin; rheumatoid arthritis;
KW diabetes; organ transplant; graft versus host disease;
KW immunosuppressant.
XX Bos taurus.
XX USS498597-A..
XX 12-MAR-1996.
PD
XX

PF 17-JAN-1992; 92US-0822966.
 PR 17-JAN-1992; 92US-0822966.
 XX
 PA (DAND) DANA FARRER CANCER INST INC.
 PA (HARD) HARVARD COLLEGE.
 PI Bierer BE, Burakoff SJ, Schreiber SL;
 XX WPI; 1996-159713/16.
 DR
 XX Purified mammalian FKBP-13 polypeptide capable of binding FK506
 PT useful for identifying and studying immunosuppressant drugs
 PT
 XX Disclosure; Column 8; 12pp; English.
 XX
 CC This sequence encoding the bovine FKBP-13 N-terminal sequence
 CC corresponds to the N-terminal sequence of human FKBP-13. FKBP-13
 CC may be used for identifying immunosuppressant drugs, and may be
 CC used in combination with immunosuppressant drugs for therapeutic
 CC purposes in the treatment of autoimmune diseases e.g. rheumatoid
 CC arthritis and type-I diabetes, organ transplant and graft versus
 CC host disease. The recombinant form of the protein could be
 CC potentially smaller and therefore easier to introduce into cells
 CC than intact FKBP-13.
 XX
 SQ Sequence 40 AA;
 Query Match 91.9%; Score 34; DB 17; Length 40;
 Best Local Similarity 71.4%; Pred. No. 6.7;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIXHC 7
 DB 14 KKRVDHC 20
 III: II
 RESULT 8
 AAG76114
 ID AAG76114 standard; Protein; 83 AA.
 AC AAG76114;
 XX
 XX 03-SEP-2001 (first entry)
 DT
 XX Human colon cancer antigen protein SEQ ID NO:6878.
 DE
 XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW colorectal carcinoma; chromosome 11.
 KW Homo sapiens.
 OS
 XX WO200122920-A2.
 PN
 XX 05-APR-2001.
 PD
 XX 28-SEP-2000; 2000WO-US26524.
 PF
 XX 29-SEP-1999; 99US-0157137.
 PR
 XX 03-NOV-1999; 99US-0163280.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;
 PI
 XX WPI; 2001-235357/24.
 DR
 XX N-PSDB; AAH35519.
 XX
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 PT
 XX Claim 11; Page 8320-8322; 9803pp; English.
 XX

CC AAH32943 to AAH37195 and AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patients own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 83 AA;
 Query Match 91.9%; Score 34; DB 22; Length 83;
 Best Local Similarity 71.4%; Pred. No. 13;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIXHC 7
 DB 48 KKRVDHC 54
 III: II
 RESULT 9
 AAR28979
 ID AAR28979 standard; protein; 99 AA.
 XX
 AC AAR28979;
 XX
 XX 25-MAR-2003 (updated)
 DT
 XX 24-MAR-1993 (first entry)
 DT
 XX Bovine RFBKP.
 DE
 XX Rapamycin; FK506; binding protein; RFBKP; prollyl isomerase;
 KW immunosuppressant; cyclosporin A; macrolide; bovine; thymus; brFKBP;
 KW cis-trans prollyl isomerase activity; FKBP12.
 XX
 OS Bos taurus.
 XX
 XX WO9219745-A1.
 PN
 XX 12-NOV-1992.
 PD
 XX 07-MAY-1992; 92WO-US03993.
 PF
 XX 08-MAY-1991; 91US-0697113.
 PR
 XX (VERT-) VERTEX PHARM INC.
 PA
 XX Harding MW;
 PI
 XX WPI; 1992-398871/48.
 DR
 XX New prollyl isomerase and rapamycin FK506 binding protein - useful
 PT for screening potential immunosuppressive cpds.
 PT
 XX Disclosure; Fig 1; 30pp; English.
 XX
 CC This sequence corresponds to a fragment of a rapamycin FK506 binding
 CC protein (RFBKP). RFBKP is a prollyl isomerase structurally related to
 CC FK506 which does not bind the immunosuppressive cyclosporin A. RFBKP
 CC binds FK506 and rapamycin with quantitatively significant selectivity.
 CC RFBKP may be used in screening assays to detect new immunosuppressants
 CC and to differentiate rapamycin-like cpds. from FK506-like cpds.

CC Rapamycin is a macrolide which is structurally related to FK506.
 CC This FKBP has been isolated from bovine thymus (bRFXBP) and was found
 CC to be of low molecular weight, approx. 16,000, and to have cis-trans
 CC prolyl isomerase activity. The N terminal of bRFXBP has been shown to
 CC have over 50% homology to the N terminal of FKBP12.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX

SQ Sequence 99 AA;

Query Match 91.9%; Score 34; DB 13; Length 99;
 Best Local Similarity 71.4%; Pred. No. 16;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KKRIXHC 7
 |||: ||
 Db 14 KKRVDHC 20

RESULT 10

AAG03758
 ID AAG03758 standard; Protein: 104 AA.

XX AC AAG03758;

XX DT 06-OCT-2000 (first entry)
 XX DE Human secreted protein, SEQ ID NO: 7839.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

DR N-PSDB; AAC03764.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX PS Claim 13; SEQ ID 7839; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX SQ Sequence 104 AA;

Query Match 91.9%; Score 34; DB 21; Length 104;
 Best Local Similarity 71.4%; Pred. No. 16;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KKRIXHC 7
 |||: ||
 Db 36 KKRVDHC 42

RESULT 11

AAB87656
 ID AAB87656 standard; protein; 137 AA.

XX AC AAB87656;

XX DT 15-MAY-2001 (first entry)

XX DE Bovine mammary tissue derived protein #47.

XX KW Bovine; mammary gland; cancer; tumour; angiogenesis.

XX OS Bos taurus.

XX PN WO200114553-A1.

XX PD 01-MAR-2001.

XX PF 23-AUG-2000; 2000WO-N200166.

XX PR 23-AUG-1999; 99US-0150330.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

XX PI (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.

XX PI Havukkala IJ, Gleen M, Grigor MR, Molenaar AJ;

XX DR WPI; 2001-226619/23.

XX PT New polypeptides and polynucleotides encoding the polypeptides, which
 PT are expressed in bovine mammary gland tissue, useful for stimulating
 PT mammary gland growth or function, or inducing differentiation of milk
 PT producing cells -

XX PS Claim 11; Page 80; 97pp; English.

XX The present invention relates to proteins derived from bovine
 CC mammary gland cells. The invention is useful for stimulating
 CC bovine mammary gland cell growth and function, inhibiting the
 CC growth of various mammary gland cancer cells, inhibiting
 CC angiogenesis and vascularization of tumours, or modulating
 CC the growth of blood vessels in a mammal.

XX SQ Sequence 137 AA;

Query Match 91.9%; Score 34; DB 22; Length 137;
 Best Local Similarity 71.4%; Pred. No. 21;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KKRIXHC 7
 |||: ||
 Db 43 KKRVDHC 49

RESULT 12

AAR93551
 ID AAR93551 standard; Protein: 141 AA.

XX AC AAR93551;

XX DT 25-JUN-1996 (first entry)

XX DE Human FKBP-13 immunophilin protein.

XX FKBP-13; immunophilin; FK506; rapamycin; rheumatoid arthritis;
 KW diabetes; organ transplant; graft versus host disease;
 KW immunosuppressant.

XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= sig_peptide
 FT Misc-difference 22..61
 FT /note= "corresponds to bovine thymus FKBP-13 N-terminal sequence"
 FT Misc-difference 117..120
 FT /note= "endoplasmic reticulum retention sequence"
 XX US5498597-A.
 XX 12-MAR-1996.
 XX 17-JAN-1992; 92US-0822966.
 XX 17-JAN-1992; 92US-0822966.
 XX (DANA) DANA FARBER CANCER INST INC.
 XX (HARD) HARVARD COLLEGE.
 XX Bierer BE, Burakoff SJ, Schreiber SL;
 XX WPI; 1996-159713/16.
 XX N-PSDB; AAT18037.
 XX Purified mammalian FKBP-13 polypeptide capable of binding FK506 -
 XX useful for identifying and studying immunosuppressant drugs
 XX Claim 1; Fig.1; 12pp; English.
 XX The FKBP-13 protein may be used for identifying immunosuppressant
 XX drugs, and may be used in combination with immunosuppressant drugs
 XX for therapeutic purposes in the treatment of autoimmune diseases e.g.
 XX rheumatoid arthritis and type-I diabetes, organ transplant and
 XX graft versus host disease. The recombinant form of the protein
 XX could be potentially smaller and therefore easier to introduce
 XX into cells than intact FKBP-13.
 XX SQ Sequence 141 AA;
 Query Match 91.9%; Score 34; DB 17; Length 141;
 Best Local Similarity 71.4%; Pred. No. 22;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIXHC 7
 DB 35 KKRVDHC 41
 RESULT 13
 ABP75992
 ID ABP75992 standard; Protein: 142 AA.
 AC ABP75992;
 XX 21-FEB-2003 (first entry)
 DE Human GENSET protein SEQ ID 199.
 XX Cytostatic; antiinflammatory; neurotropic; neuroprotective; cardiatic;
 XX gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;
 KW inflammatory disease; immune disorder; neuromuscular; toxicity;
 KW central nervous system; cardiovascular; gastrointestinal.
 XX Homo sapiens.
 XX WO200283898-A1.
 XX 24-OCT-2002.

PF 18-APR-2001; 2001WO-IB00914.
 XX 18-APR-2001; 2001WO-IB00914.
 XX (GEST) GENSET.
 XX Bejanin S, Tanaka H, Dumas Milne Edwards J, Jobert S, Giordano J;
 XX WPI; 2003-075548/07.
 XX New GENSET polynucleotides and polypeptides, useful for treating heavy
 XX metal toxicity, cancer, inflammatory diseases, immune disorders, and
 XX the neuromuscular, CNS, cardiovascular or gastrointestinal effects of
 XX the toxicity
 XX Claim 14; Page 414; 735pp; English.
 XX The present invention relates to novel GENSET polynucleotides
 XX (AB236404-AB236911) encoding polypeptides (ABP75963-ABP76368). The
 XX polynucleotides and polypeptides are useful in screening and diagnostic
 XX assays for abnormal GENSET expression and/or biological activity. They
 XX are also useful for screening of compounds for treating or preventing
 XX GENSET-related disorders, such as heavy metal toxicity, cancer,
 XX inflammatory diseases, immune disorders, and the neuromuscular, central
 XX nervous system (CNS), cardiovascular or gastrointestinal effects of the
 XX toxicity.
 XX SQ Sequence 142 AA;
 Query Match 91.9%; Score 34; DB 24; Length 142;
 Best Local Similarity 71.4%; Pred. No. 22;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIXHC 7
 DB 36 KKRVDHC 42
 RESULT 14
 ABP76156
 ID ABP76156 standard; Protein: 142 AA.
 XX AC ABP76156;
 XX 21-FEB-2003 (first entry)
 DE Human GENSET protein SEQ ID 482.
 XX Cytostatic; antiinflammatory; neurotropic; neuroprotective; cardiatic;
 KW gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;
 KW inflammatory disease; immune disorder; neuromuscular; toxicity;
 KW central nervous system; cardiovascular; gastrointestinal.
 XX Homo sapiens.
 XX WO200283898-A1.
 XX 24-OCT-2002.
 XX 18-APR-2001; 2001WO-IB00914.
 XX 18-APR-2001; 2001WO-IB00914.
 XX (GEST) GENSET.
 XX Bejanin S, Tanaka H, Dumas Milne Edwards J, Jobert S, Giordano J;
 XX WPI; 2003-075548/07.
 XX New GENSET polynucleotides and polypeptides, useful for treating heavy
 XX metal toxicity, cancer, inflammatory diseases, immune disorders, and
 XX the neuromuscular, CNS, cardiovascular or gastrointestinal effects of
 XX the toxicity

XX PS Claim 14; Page 545; 735pp; English.

XX CC The present invention relates to novel GENSET polynucleotides

XX CC (AB236404-AB236911) encoding polypeptides (ABP75963-ABP76368). The

XX CC polynucleotides and polypeptides are useful in screening and diagnostic

XX CC assays for abnormal GENSET expression and/or biological activity. They

XX CC are also useful for screening of compounds for treating or preventing

XX CC GENSET-related disorders, such as heavy metal toxicity, cancer,

XX CC inflammatory diseases, immune disorders, and the neuromuscular, central

XX CC nervous system (CNS), cardiovascular or gastrointestinal effects of the

XX CC toxicity.

XX SQ Sequence 142 AA;

Query Match 91.9%; Score 34; DB 24; Length 142;

Best Local Similarity 71.4%; Pred. No. 22;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1'KKRIXHC 7

Db 36 KKRVDHC 42

RESULT 15

ABG60082

ID ABG60082 standard; Protein; 163 AA.

AC ABG60082;

XX 30-JUL-2002 (first entry) .

XX Human DITHP polypeptide #140.

XX Human; DITHP: diagnostic and therapeutic polypeptide; bone; testis; skin;

XX cell proliferative disorder; cancer; tumour; autoimmune disorder; brain;

XX inflammatory disorder; viral infection; bacterial infection; seizure;

XX fungal infection; parasitic infections; developmental disorder; breast;

XX endocrine disorder; metabolic disorder; neurological disorder; cervix;

XX gastrointestinal disorder; transport disorder; gene therapy; kidney;

XX adrenal gland; bone marrow; lung; ovary; pancreas; prostate; spleen;

XX thymus.

OS Homo sapiens.

XX WO200220754-A2.

XX 14-MAR-2002.

XX 29-AUG-2001; 2001WO-US27127.

XX 05-SEP-2000; 2000US-229747P.

XX 05-SEP-2000; 2000US-229748P.

XX 05-SEP-2000; 2000US-229749P.

XX 05-SEP-2000; 2000US-229750P.

XX 05-SEP-2000; 2000US-229751P.

XX 05-SEP-2000; 2000US-230583P.

XX 06-SEP-2000; 2000US-230505P.

XX 06-SEP-2000; 2000US-230514P.

XX 06-SEP-2000; 2000US-230515P.

XX 06-SEP-2000; 2000US-230517P.

XX 06-SEP-2000; 2000US-230518P.

XX 06-SEP-2000; 2000US-230519P.

XX 06-SEP-2000; 2000US-230595P.

XX 06-SEP-2000; 2000US-230597P.

XX 06-SEP-2000; 2000US-230598P.

XX 06-SEP-2000; 2000US-230599P.

XX 06-SEP-2000; 2000US-230610P.

XX 06-SEP-2000; 2000US-230865P.

XX 06-SEP-2000; 2000US-230988P.

XX 07-SEP-2000; 2000US-230951P.

XX 07-SEP-2000; 2000US-231163P.

XX 07-SEP-2000; 2000US-231167P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;

XX PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;

XX PI Momiya MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;

XX PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;

XX PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;

XX WPI: 2002-383054/41.

XX DR N-PSDB; ABK71673.

XX PT An isolated polynucleotide useful in diagnostics and therapeutics -

XX PS Claim 29; Page 613-614; 686pp; English.

XX CC The invention relates to human diagnostic and therapeutic (dithp)

XX CC polynucleotides and their associated polypeptides (DITHP polypeptides).

XX CC The sequences of the invention are used in the treatment and diagnosis of

XX CC cell proliferative disorders (e.g. atherosclerosis, cirrhosis), cancers

XX CC (e.g. tumours of the adrenal gland, bone, bone marrow, brain, breast,

XX CC cervix, kidney, lung, ovary, pancreas, prostate, skin, spleen, testis or

XX CC thymus), autoimmune/inflammatory disorders (e.g. asthma, bronchitis,

XX CC psoriasis, osteoporosis), viral infections, bacterial infections, fungal

XX CC infections, parasitic infections, developmental disorders (e.g. anaemia,

XX CC epilepsy), seizure disorders (e.g. cerebral palsy, spina bifida),

XX CC endocrine disorders (e.g. thrombosis, aneurysm), metabolic disorders

XX CC (e.g. obesity, diabetes), neurological disorders (e.g. stroke,

XX CC amyotrophic lateral sclerosis, multiple sclerosis), gastrointestinal

XX CC disorders (e.g. ulcerative colitis, lysinuria) and transport disorders

XX CC (e.g. myotonic dystrophy, catatonias, peripheral neuropathy). Sequences

XX CC ABG59943-ABG60220 represent human DITHP polypeptides of the invention.

XX SQ Sequence 163 AA;

Query Match 91.9%; Score 34; DB 23; Length 163;

Best Local Similarity 71.4%; Pred. No. 25;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1'KKRIXHC 7

Db 57 KKRVDHC 63

Search completed: July 29, 2003, 09:54:55

Job time : 35.0667 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:53:43 ; Search time 13.5333 Seconds
(without alignments)
21.885 Million cell updates/sec

Title: US-09-606-129A-16
Perfect score: 37
Sequence: 1 KKRIHXC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	91.9	40	1 US-07-822-966B-4	Sequence 4, Appli
2	34	91.9	88	1 US-08-336-618-19	Sequence 19, Appli
3	34	91.9	99	5 PCT-US92-03993-5	Sequence 5, Appli
4	34	91.9	141	1 US-07-822-966B-6	Sequence 6, Appli
5	34	91.9	141	2 US-08-803-899-6	Sequence 6, Appli
6	34	91.9	142	1 US-08-336-618-18	Sequence 18, Appli
7	34	91.9	142	5 PCT-US92-03993-7	Sequence 7, Appli
8	30	81.1	504	3 US-09-231-529-6	Sequence 6, Appli
9	30	81.1	504	3 US-08-977-816-6	Sequence 6, Appli
10	30	81.1	527	4 US-09-659-166-2	Sequence 2, Appli
11	30	81.1	528	2 US-08-802-466-2	Sequence 2, Appli
12	30	81.1	528	3 US-09-350-484-2	Sequence 2, Appli
13	30	81.1	565	2 US-08-892-770-5	Sequence 6, Appli
14	30	81.1	571	2 US-08-892-770-5	Sequence 5, Appli
15	29	78.4	269	3 US-09-006-783A-5	Sequence 5, Appli
16	29	78.4	269	4 US-09-159-871-11	Sequence 11, Appli
17	29	78.4	294	2 US-09-258-371-10	Sequence 10, Appli
18	29	78.4	294	3 US-08-751-230-10	Sequence 10, Appli
19	29	78.4	294	3 US-09-499-082-10	Sequence 10, Appli
20	29	78.4	294	3 US-09-258-372-10	Sequence 10, Appli
21	29	78.4	294	3 US-09-006-783A-3	Sequence 3, Appli
22	29	78.4	294	4 US-09-159-871-2	Sequence 2, Appli
23	29	78.4	477	2 US-08-560-098A-51	Sequence 51, Appli
24	29	78.4	521	4 US-09-252-991A-25430	Sequence 25430, A
25	28	75.7	22	1 US-08-484-635-166	Sequence 166, App
26	28	75.7	22	2 US-08-484-631-166	Sequence 166, App
27	28	75.7	22	2 US-08-827-570-166	Sequence 166, App

28	75.7	189	4	US-09-107-532A-4526	Sequence 4526, Ap
29	75.7	398	4	US-09-328-352-4507	Sequence 4507, Ap
30	75.7	508	2	US-08-818-024-3	Sequence 3, Appli
31	75.7	508	3	US-09-334-775A-3	Sequence 3, Appli
32	75.7	513	3	US-09-041-886-28	Sequence 28, Appli
33	75.7	530	3	US-09-041-886-29	Sequence 29, Appli
34	75.7	552	3	US-09-041-886-30	Sequence 30, Appli
35	75.7	589	3	US-09-041-886-31	Sequence 31, Appli
36	75.7	717	3	US-08-872-855-9	Sequence 9, Appli
37	75.7	1058	4	US-09-252-991A-29105	Sequence 29105, A
38	75.7	3118	2	US-08-457-273B-8	Sequence 8, Appli
39	75.7	3119	1	US-08-246-982A-16	Sequence 16, Appli
40	75.7	3119	1	US-08-453-265-16	Sequence 16, Appli
41	75.7	3144	1	US-08-246-982A-6	Sequence 6, Appli
42	75.7	3144	1	US-08-453-265-6	Sequence 6, Appli
43	75.7	3144	2	US-08-457-273B-42	Sequence 42, Appli
44	75.7	3144	3	US-08-556-419-21	Sequence 21, Appli
45	75.7	3144	3	US-09-041-886-15	Sequence 15, Appli

ALIGNMENTS

RESULT 1
US-07-822-966B-4
; Sequence 4, Application US/07822966B
; Patent No. 5498597
; GENERAL INFORMATION:
; APPLICANT: Steven J. Burakoff
; APPLICANT: Stuart L. Schreiber
; APPLICANT: Barbara E. Bierer
; TITLE OF INVENTION: FKBP-13, AN FK506-BINDING
; TITLE OF INVENTION: IMMUNOPHILIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50z or 55sx
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/822,966B
; FILING DATE: January 17, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/052001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-07-822-966B-4

Query Match 91.9%; Score 34; DB 1; Length 40;
Best Local Similarity 71.4%; Pred. NO. 1.2;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 KKRIHXC 7
Db 14 KKRYDHC 20

RESULT 2
US-08-336-618-19
; Sequence 19, Application US/08336618
; Patent No. 5763590
; GENERAL INFORMATION:
; APPLICANT: Peattie, Debra A.
; APPLICANT: Harding, Matthew W.
; APPLICANT: Livingston, David J.
; TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
; TITLE OF INVENTION: PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
; TITLE OF INVENTION: CDNA
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,618
; FILING DATE: 09-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/963,325
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/777,752
; FILING DATE: 11-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/
; FILING DATE: 09-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: VPI91-06A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-618-19
Query Match 91.9%; Score 34; DB 1; Length 88;
Best Local Similarity 71.4%; Pred. No. 2.5;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KKRXHC 7
Db 14 KKRVDHC 20
RESULT 3
PCT-US92-03993-5
; Sequence 5, Application PC/TUS9203993
; GENERAL INFORMATION:
; APPLICANT: Harding, Matthew W.
; TITLE OF INVENTION: FKBP: A NOVEL PROLYL ISOMERASE AND
; TITLE OF INVENTION: RAPAMYCIN/FK506 BINDING PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; OPERATING SYSTEM: IBM PS/2 Model 502 or 55SX
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/822,966B
; FILING DATE: January 17, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/052001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154

STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/03993
FILING DATE: 19920507
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/697,113
FILING DATE: 08-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: VPI91-05A PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 99 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-03993-5
Query Match 91.9%; Score 34; DB 5; Length 99;
Best Local Similarity 71.4%; Pred. No. 2.8;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 KKRXHC 7
Db 14 KKRVDHC 20
RESULT 4
US-07-822-966B-6
; Sequence 6, Application US/07822966B
; Patent No. 5498597
; GENERAL INFORMATION:
; APPLICANT: Steven J. Burakoff
; APPLICANT: Stuart L. Schreiber
; APPLICANT: Barbara E. Bierer
; TITLE OF INVENTION: FKBP-13, AN FK506-BINDING
; TITLE OF INVENTION: IMMUNOPHILIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; OPERATING SYSTEM: IBM PS/2 Model 502 or 55SX
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/822,966B
; FILING DATE: January 17, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/052001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-822-966B-6

Query Match 91.9%; Score 34; DB 1; Length 141;
Best Local Similarity 71.4%; Pred. No. 3.8;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKR1XHC 7
|||: ||
Db 35 KKRVDHC 41

RESULT 5

US-08-803-899-6
Sequence 6, Application US/08803899
Patent No. 5912224
GENERAL INFORMATION:
APPLICANT: DONAHOE, PATRICIA K.
APPLICANT: WANG, TONGWEN
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
CELLULAR RESPONSE TO TGF-BETA LIGANDS
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/803.899
FILING DATE: 02/21/1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/012,054
FILING DATE: 02/22/1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36-688
REFERENCE/DOCKET NUMBER: 0609.4240001

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-803-899-6

Query Match 91.9%; Score 34; DB 2; Length 141;
Best Local Similarity 71.4%; Pred. No. 3.8;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKR1XHC 7
|||: ||
Db 35 KKRVDHC 41

RESULT 6

US-08-336-618-18
Sequence 18, Application US/08336618
Patent No. 5763590
GENERAL INFORMATION:
APPLICANT: Peattie, Debra A.
APPLICANT: Harding, Matthew W.
APPLICANT: Livingston, David J.
TITLE OF INVENTION: ISOLATION OF AN MR 52,000 FK506 BINDING
PROTEIN AND MOLECULAR CLONING OF A CORRESPONDING HUMAN
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith and Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336.618
FILING DATE: 09-NOV-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/963,325
FILING DATE: 16-OCT-1992
APPLICATION NUMBER: US 07/777,752
FILING DATE: 11-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/
FILING DATE: 09-OCT-1992
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: VPI91-06A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 142 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-618-18

Query Match 91.9%; Score 34; DB 1; Length 142;
Best Local Similarity 71.4%; Pred. No. 3.8;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKR1XHC 7
|||: ||
Db 36 KKRVDHC 42

RESULT 7

PCT-US92-03993-7
Sequence 7, Application PC/TUS9203993
GENERAL INFORMATION:
APPLICANT: Harding, Matthew W.
TITLE OF INVENTION: FKBP: A NOVEL PROLYL ISOMERASE AND
NUMBER OF SEQUENCES: 10
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA

COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/03993
FILING DATE: 19920507
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/697,113
FILING DATE: 08-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: VPI91-05A PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 7:
LENGTH: 142 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-03993-7

Query Match 91.1%; Score 34; DB 5; Length 142;
Best Local Similarity 71.4%; Pred. No. 3.8;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRXXHC 7
Db 36 KKRVDHC 42

RESULT 8
US-09-231-529-6
Sequence 6, Application US/09231529
Patent No. 6096308
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEIN KINASE AND KINASE INHIBITORS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/231,529
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/977,816
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0429 US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 504 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT11
CLONE: 701698
US-09-231-529-6

Query Match 81.1%; Score 30; DB 3; Length 504;
Best Local Similarity 71.4%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRXXHC 7
Db 309 KERICHC 315

RESULT 9
US-08-977-816-6
Sequence 6, Application US/08977816
Patent No. 6194186
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN PROTEIN KINASE AND KINASE INHIBITORS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,816
FILING DATE: Filed Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0429 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 504 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNNOT11
CLONE: 701698
US-08-977-816-6

Query Match 81.1%; Score 30; DB 3; Length 504;

Best Local Similarity 71.4%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIHXC 7
|:|:| |
Db 309 KERICH 315

RESULT 10

US-09-659-166-2
; Sequence 2, Application US/09659166
; Patent No. 6355465
; GENERAL INFORMATION:
; APPLICANT: CREASY, CARETHA LEE
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30218
; CURRENT APPLICATION NUMBER: US/09/659,166
; CURRENT FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: UK 9921505.5
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 527
; TYPE: PRT
; ORGANISM: rattus
US-09-659-166-2

Query Match 81.1%; Score 30; DB 4; Length 527;
Best Local Similarity 71.4%; Pred. No. 88;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KKRIHXC 7
|:|:| |
Db 267 KNRIHC 273

RESULT 11

US-08-802-466-2
; Sequence 2, Application US/08802466
; Patent No. 5972606
; GENERAL INFORMATION:
; APPLICANT: Creasy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,466
; FILING DATE: 19 February 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX:

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 528 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-802-466-2

Query Match 81.1%; Score 30; DB 2; Length 528;
Best Local Similarity 71.4%; Pred. No. 88;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KKRIHXC 7
|:|:| |
Db 268 KNRIHC 274

RESULT 12

US-09-350-484-2
; Sequence 2, Application US/09350484
; Patent No. 6159716
; GENERAL INFORMATION:
; APPLICANT: Creasy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/350,484
; FILING DATE: 09-Jul-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/802,466
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GH50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-4026
; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 528 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-350-484-2

Query Match 81.1%; Score 30; DB 3; Length 528;
Best Local Similarity 71.4%; Pred. No. 88;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KKRIHXC 7
|:|:| |
Db 268 KNRIHC 274

RESULT 13

US-08-892-770-6
; Sequence 6, Application US/08892770
; Patent No. 5962670
; GENERAL INFORMATION:
; APPLICANT: Walling, Linda L.
; APPLICANT: Pautot, Veronique
; APPLICANT: Gu, Yong-Qiang
; APPLICANT: Chao, Wun Shaw
; TITLE OF INVENTION: Improved Promoters for Enhancing Plant
; PRODUCTIVITY
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/892,770
; FILING DATE: 15-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 023070-0721000S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 565 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..565
; OTHER INFORMATION: /note= "deduced partial protein sequence
; OF ACIDIC LEUCINE AMINOPEPTIDASE 2
; OTHER INFORMATION: (LapA2) from tomato"
US-08-892-770-6
Query Match 81.1%; Score 30; DB 2; Length 565;
Best Local Similarity 83.3%; Pred. No. 94;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 KRXXHC 7
Db 42 KRIVHC 47
RESULT 14
US-08-892-770-5
; Sequence 5, Application US/08892770
; Patent No. 5962670
; GENERAL INFORMATION:
; APPLICANT: Walling, Linda L.
; APPLICANT: Pautot, Veronique
; APPLICANT: Gu, Yong-Qiang
; APPLICANT: Chao, Wun Shaw
; TITLE OF INVENTION: Improved Promoters for Enhancing Plant
; PRODUCTIVITY
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/892,770
FILING DATE: 15-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 023070-0721000S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 571 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..571
OTHER INFORMATION: /note= "deduced full-length protein
sequence of acidic leucine
aminopeptidase (LapA1) protein from
tomato"
US-08-892-770-5
Query Match 81.1%; Score 30; DB 2; Length 571;
Best Local Similarity 83.3%; Pred. No. 95;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 KRXXHC 7
Db 48 KRIVHC 53
RESULT 15
US-09-006-783A-5
; Sequence 5, Application US/09006783A
; Patent No. 6297366
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrey V
; APPLICANT: Garkavstev, Igor
; APPLICANT: Riabowol, Karl
; TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling
; PATHWAY
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/006,783A
; FILING DATE: 15-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

NAME: NO. 6297366nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 97,837
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 269 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-006-783A-5

Query Match 78.4%; Score 29; DB 3; Length 269;
Best Local Similarity 57.1%; Pred No. 76;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRXXHC 7
|:|:|
Db 67 KRRMLHC 73

Search completed: July 29, 2003, 09:57:30
Job time : 15.5333 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:55:03 ; Search time 16.8 seconds
(without alignments)
49.483 Million cell updates/sec

Title: US-09-606-129A-16
Perfect score: 37
Sequence: 1 KKRXXHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	94.6	7	15	US-10-045-545-16
2	35	94.6	295	15	US-10-045-545-4
3	35	94.6	296	15	US-10-045-545-1
4	35	94.6	296	15	US-10-045-545-3
5	34	91.9	83	15	US-10-106-698-6888
6	30	81.1	141	9	US-09-925-299-787
7	30	81.1	141	11	US-09-925-299-787
8	30	81.1	526	9	US-09-855-154-2
9	30	81.1	528	15	US-10-253-349-6
10	29	78.4	63	10	US-09-764-877-1283
11	29	78.4	91	9	US-09-864-761-36953
12	29	78.4	175	10	US-09-764-864-1357
13	29	78.4	179	10	US-09-764-864-913
14	29	78.4	269	11	US-09-968-653A-5
15	29	78.4	294	11	US-09-968-653A-3

16	28	75.7	86	11	US-09-764-891-4720	Sequence 4720, Ap
17	28	75.7	372	9	US-09-741-669-367	Sequence 367, App
18	28	75.7	372	9	US-09-912-020-375	Sequence 375, App
19	28	75.7	712	15	US-10-128-714-3108	Sequence 3108, Ap
20	28	75.7	716	15	US-10-128-714-8108	Sequence 8108, Ap
21	28	75.7	828	10	US-09-738-626-5038	Sequence 5038, Ap
22	28	75.7	1543	9	US-09-904-987-7	Sequence 7, Appli
23	27	73.0	150	10	US-09-731-872-333	Sequence 333, App
24	27	73.0	158	15	US-10-101-464A-705	Sequence 705, App
25	27	73.0	177	10	US-09-731-872-360	Sequence 360, App
26	27	73.0	396	14	US-10-155-613-2	Sequence 2, Appli
27	27	73.0	406	9	US-09-815-242-5598	Sequence 5598, Ap
28	27	73.0	407	9	US-09-925-301-1057	Sequence 1057, Ap
29	27	73.0	422	9	US-09-815-242-12287	Sequence 12287, A
30	27	73.0	457	9	US-09-815-242-11616	Sequence 11616, A
31	27	73.0	492	15	US-10-007-280A-212	Sequence 212, App
32	27	73.0	501	15	US-10-106-698-4629	Sequence 4629, Ap
33	27	73.0	551	15	US-10-253-349-3	Sequence 3, Appli
34	27	73.0	587	15	US-10-253-349-2	Sequence 2, Appli
35	27	73.0	597	15	US-10-025-222A-40	Sequence 40, Appl
36	27	73.0	873	9	US-09-954-043-2	Sequence 2, Appli
37	27	73.0	873	15	US-10-200-154-2	Sequence 2, Appli
38	26	70.3	37	11	US-09-764-891-3816	Sequence 3816, Ap
39	26	70.3	81	10	US-09-738-626-5957	Sequence 5957, Ap
40	26	70.3	81	15	US-10-106-698-6169	Sequence 6169, Ap
41	26	70.3	145	15	US-10-029-359A-4	Sequence 4, Appli
42	26	70.3	156	15	US-10-043-487-412	Sequence 412, App
43	26	70.3	167	11	US-09-791-279-186	Sequence 186, App
44	26	70.3	170	11	US-09-934-455-464	Sequence 464, App
45	26	70.3	176	15	US-10-043-487-401	Sequence 401, App

ALIGNMENTS

RESULT 1
US-10-045-545-16
; Sequence 16, Application US/10045545
; Publication No. US20030027124A1
; GENERAL INFORMATION:
; APPLICANT: Maines, Mahin D.
; TITLE OF INVENTION: METHODS OF MODIFYING CELL STRUCTURE AND REMODELING
; FILE REFERENCE: 176/60981
; CURRENT APPLICATION NUMBER: US/10/045,545
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/261,500
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
; OTHER INFORMATION: kinase C enhancing domain
; NAME/KEY: PEPTIDE
; LOCATION: (5)
; OTHER INFORMATION: where X is any aa
US-10-045-545-16

Query Match 94.6% Score 35; DB 15; Length 7;
Best Local Similarity 100.0%; Pred. No. 4e+05; 0; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 1 KKRXXHC 7
Db 1 KKRXXHC 7

RESULT 2
US-10-045-545-4

; Sequence 4, Application US/10045545
; Publication No. US20030027124A1
; GENERAL INFORMATION:
; APPLICANT: Maines, Mahin D.
; TITLE OF INVENTION: METHODS OF MODIFYING CELL STRUCTURE AND REMODELING
; FILE REFERENCE: 176/60981
; CURRENT APPLICATION NUMBER: US/10/045,545
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: 60/261,500
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-045-545-4

Query Match 94.6%; Score 35; DB 15; Length 295;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRXXHC 7
DB 274 KKRIMHC 280

RESULT 3

US-10-045-545-1
; Sequence 1, Application US/10045545
; Publication No. US20030027124A1
; GENERAL INFORMATION:
; APPLICANT: Maines, Mahin D.
; TITLE OF INVENTION: METHODS OF MODIFYING CELL STRUCTURE AND REMODELING
; FILE REFERENCE: 176/60981
; CURRENT APPLICATION NUMBER: US/10/045,545
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: 60/261,500
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 1
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-045-545-1

Query Match 94.6%; Score 35; DB 15; Length 296;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRXXHC 7
DB 275 KKRILHC 281

RESULT 4

US-10-045-545-3
; Sequence 3, Application US/10045545
; Publication No. US20030027124A1
; GENERAL INFORMATION:
; APPLICANT: Maines, Mahin D.
; TITLE OF INVENTION: METHODS OF MODIFYING CELL STRUCTURE AND REMODELING
; FILE REFERENCE: 176/60981
; CURRENT APPLICATION NUMBER: US/10/045,545
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: 60/261,500
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-045-545-3

Query Match 94.6%; Score 35; DB 15; Length 296;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRXXHC 7
DB 275 KKRILHC 281

RESULT 5

US-10-106-698-6888
; Sequence 6888, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypept
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6888
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (74)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (77)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6888

Query Match 91.9%; Score 34; DB 15; Length 83;
Best Local Similarity 71.4%; Pred. No. 9.2;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRXXHC 7
DB 48 KKRVDHC 54

RESULT 6

US-09-925-299-787
; Sequence 787, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 787
; LENGTH: 141

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-787

Query Match 81.1%; Score 30; DB 9; Length 141;
Best Local Similarity 71.4%; Pred. No. 87;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KKRIXHC 7
Db 27 KNRIIHC 33

RESULT 7

US-09-925-299-787
; Sequence 787, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 787
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-787

Query Match 81.1%; Score 30; DB 11; Length 141;
Best Local Similarity 71.4%; Pred. No. 87;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KKRIXHC 7
Db 27 KNRIIHC 33

RESULT 8

US-09-855-154-2
; Sequence 2, Application US/09855154
; Patent No. US20020064852A1
; GENERAL INFORMATION:
; APPLICANT: CARETHA L. BURNS
; TITLE OF INVENTION: MURINE SERINE/THREONINE KINASE, mDYRK2
; FILE REFERENCE: GH-70699
; CURRENT APPLICATION NUMBER: US/09/855,154
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/204,489
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 526
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-855-154-2

Query Match 81.1%; Score 30; DB 9; Length 526;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KKRIXHC 7
Db 266 KNRIIHC 272

RESULT 9

US-10-253-349-6
; Sequence 6, Application US/10253349
; Publication No. US20030077679A1
; GENERAL INFORMATION:
; APPLICANT: ANNAN, ROLAND S.
; APPLICANT: BURNS, BRIAN M.
; APPLICANT: CREASY, CARETHA LEE
; APPLICANT: HEROLD, KATHLEEN M.
; TITLE OF INVENTION: METHODS FOR FINDING MODULATORS OF DYRK3 AND DYRK2
; FILE REFERENCE: GH50044.
; CURRENT APPLICATION NUMBER: US/10/253,349
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: 60/326,973
; PRIOR FILING DATE: 2001-10-04
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 528
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-253-349-6

Query Match 81.1%; Score 30; DB 15; Length 528;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KKRIXHC 7
Db 268 KNRIIHC 274

RESULT 10

US-09-764-877-1283
; Sequence 1283, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1283
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (11)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (61)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1283

Query Match 78.4%; Score 29; DB 10; Length 63;
Best Local Similarity 42.9%; Pred. No. 69;
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIXHC 7
Db 30 RKKVTHC 36

RESULT 11

US-09-864-761-36953

Sequence 36953, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 36953
LENGTH: 91
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC012073.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.6
OTHER INFORMATION: SWISSPROT HIT: P39057, EVALUE 2.00e+00
OTHER INFORMATION: EST_HUMAN HIT: BE901209.1, EVALUE 5.00e-30
US-09-864-761-36953
Query Match 78.4%; Score 29; DB 9; Length 91;
Best Local Similarity 71.4%; Pred. No. 94;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KKRIXHC 7
Db 75 KLRISHC 81
RESULT 12
US-09-764-864-1357
Sequence 1357, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PTZ23
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1357
LENGTH: 175
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-864-1357
Query Match 78.4%; Score 29; DB 10; Length 175;
Best Local Similarity 57.1%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 KKRIXHC 7
Db 144 KHLTHC 150
RESULT 13
US-09-764-864-913
Sequence 913, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PTZ23
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 913
LENGTH: 179
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-864-913
Query Match 78.4%; Score 29; DB 10; Length 179;
Best Local Similarity 57.1%; Pred. No. 1.7e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 KKRIXHC 7
Db 148 KHLTHC 154
RESULT 14
US-09-968-653A-5
Sequence 5, Application US/09968653A
Publication No. US20030073084A1
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrey V
Riabowol, Karl
Garkavstev, Igor
TITLE OF INVENTION: p53/ING1 as a Mediator of p53 Signaling
Pathway
NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/968,653A
FILING DATE: 01-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/006,783A
FILING DATE: 15-JAN-1998
ATTORNEY/AGENT INFORMATION:
NAME: No. US20030073084Alnan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 97,837
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
LENGTH: 269 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-968-653A-5
Query Match 78.4%; Score 29; DB 11; Length 269;
Best Local Similarity 57.1%; Pred. No. 2.4e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0:
Qy 1 KKRIXHC 7
Db 67 KRRMLHC 73
RESULT 15
US-09-653A-3
Sequence 3, Application US/09968653A
Publication No. US20030073084A1
GENERAL INFORMATION:
APPLICANT: Gudkov, Andrey V
Garkavstev, Igor
Riabowol, Karl
TITLE OF INVENTION: p33/ING1 as a Mediator of p53 Signaling
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/968,653A
FILING DATE: 01-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER: US/09/006,783A

FILING DATE: 15-JAN-1998
ATTORNEY/AGENT INFORMATION:
NAME: No. US20030073084Alnan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 97,837
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-968-653A-3
Query Match 78.4%; Score 29; DB 11; Length 294;
Best Local Similarity 57.1%; Pred. No. 2.6e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0:
Qy 1 KKRIXHC 7
Db 82 KRRMLHC 88
Search completed: July 29, 2003, 09:58:13
Job time : 17.8 secs

THIS PAGE BLANK (uspro)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:53:07 ; Search time 12.1333 Seconds
(without alignments)
55.482 Million cell updates/sec

Title: US-09-606-129A-16

Perfect score: 37

Sequence: 1 KKRXHC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	94.6	295	A42268	biliverdin reducta
2	35	94.6	296	G02066	biliverdin reducta
3	35	94.6	296	S62624	biliverdin reducta
4	35	94.6	401	F85015	probable phosphati
5	35	94.6	431	T01723	1-phosphatidylinos
6	35	94.6	633	T04537	hypothetical prote
7	34	91.9	140	I49668	binding protein -
8	34	91.9	142	JC1365	FK506/rapamycin-bi
9	34	91.9	925	S50490	hypothetical prote
10	33	89.2	189	A38117	hypothetical prote
11	33	89.2	189	AD2414	transposase ali1486
12	33	89.2	189	AE2064	transposase ali1206
13	33	89.2	189	AF2052	transposase ali1197
14	33	89.2	189	AF2152	transposase ali277
15	33	89.2	189	AG1875	transposase ali055
16	33	89.2	189	AG2037	transposase ali185
17	33	89.2	512	H86206	hypothetical prote
18	33	89.2	321	T48373	hypothetical prote
19	31	83.8	457	T21344	hypothetical prote
20	30	81.1	38	E72247	ribosomal protein
21	30	81.1	108	A64511	hypothetical prote
22	30	81.1	197	E95203	conserved hypothet
23	30	81.1	197	E98070	conserved hypothet
24	30	81.1	411	C86145	hypothetical prote
25	30	81.1	463	D84065	glucose-1-phosphat
26	30	81.1	504	JC4775	interferon-induced
27	30	81.1	504	A56534	interferon-induced
28	30	81.1	508	T02486	hypothetical prote-
29	30	81.1	566	T07850	leucyl aminopeptid

ALIGNMENTS

RESULT 1

A42268

biliverdin reductase (EC 1.3.1.24) - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999

C:Accession: A42268

R:Fakhrai, H.; Maines, M.D.

J. Biol. Chem. 267, 4023-4029, 1992

A:Title: Expression and characterization of a cDNA for rat kidney biliverdin reductase

A:Reference number: A42268; MUID:92156147; PMID:1371282

A:Accession: A42268

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid; protein

A:Residues: 1-295 <FAK>

A:Cross-references: GB:M81681; NID:g203177; PIDN:AAA40830.1; PID:g203178

A:Experimental source: Kidney

A>Note: sequence extracted from NCBI backbone (NCBIP:82800)

C:Keywords: liver; oxidoreductase

Query Match 94.6%; Score 35; DB 2; Length 295;

Best Local Similarity 85.7%; Pred. No. 4.6;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRXHC 7

|||||

Db 274 KKRIMHC 280

RESULT 2

G02066

biliverdin reductase (EC 1.3.1.24) - human

N:Alternate names: biliverdin IX-alpha reductase

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998

C:Accession: G02066

R:Komuro, A.; Tobe, T.; Nakano, Y.; Yamaguchi, T.; Tomita, M.

submitted to the EMBL Data Library, August 1995

A:Reference number: H00768

A:Accession: G02066

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-296 <KOM>

A:Cross-references: EMBL:U34877; NID:g1143231; PID:g1143232

C:Keywords: Oxidoreductase

Query Match 94.6%; Score 35; DB 2; Length 296;

Best Local Similarity 85.7%; Pred. No. 4.6;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRXHC 7

|||||

Db 275 KKRILHC 281

```

RESULT 3
S62624
biliverdin reductase (EC 1.3.1.24) - human
N:Alternate names: biliverdin IX-alpha reductase
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1996 #sequence_revision 09-May-1997 #text_change 21-Jul-2000
C:Accession: S62624; S62622; S29736
R:Waines, M.D.; Polevoda, B.V.; Huang, T.J.; McCoubrey Jr., W.K.
Eur. J. Biochem. 235, 372-381, 1996
A:Title: Human biliverdin IX-alpha reductase is a zinc-metalloprotein. Characterization
A:Reference number: S62622; MUID:96202961; PMID:8631357
A:Accession: S62624
A:Molecule type: mRNA
A:Residues: 1-296 <MAI>
A:Cross-references: EMBL:X93086; NID:g1246748; PIDN:CAA63635.1; PID:g1246749
A:Accession: S62622
A:Molecule type: protein
A:Residues: 3-24, 'X', 26-27, 'X', 29-36; 48-74; 228-234; 235-248 <MAF>
R:Waines, M.D.; Trakshel, G.M.
Arch. Biochem. Biophys. 300, 320-326, 1993
A:Title: Purification and characterization of human biliverdin reductase.
A:Reference number: S29736; MUID:93143333; PMID:8424666
A:Accession: S29736
A:Molecule type: protein
A:Residues: 3-24, 'X', 26-27, 'X', 29-36; 48-74; 228-234; 235-248 <MAW>
A:Note: the sequence of peptide 1 from page 323 seems not to belong to this protein
C:Genetics:
A:Gene: BVR
C:Keywords: oxidoreductase
F:3-296/Product: biliverdin reductase IX-alpha #status experimental <MAT>

Query Match          94.6%; Score 35; DB 2; Length 296;
Best Local Similarity 85.7%; Pred. No. 4.6;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRXXHC 7
    |||||
Db 275 KKRILHC 281

RESULT 4
F85015
probable phosphatidylinositol kinase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: F85015
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: F85015
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <STO>
A:Cross-references: GB:NC_001268; NID:97267616; PIDN:CAB80928.1; GSPDB:GN00140
A:Gene: AT4G01190
A:Map position: 4

Query Match          94.6%; Score 35; DB 2; Length 401;
Best Local Similarity 85.7%; Pred. No. 6.1;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRXXHC 7
    |||||
Db 353 KKRIEHC 359

RESULT 5
T01723
1-phosphatidylinositol-4-phosphate 5-kinase type II homolog - Arabidopsis thaliana

```

```

N:Alternate names: protein A_IG002N01.9
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01723
R:Scheet, P.; Maggi, L.
Submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana IG002N01.
A:Reference number: Z14407
A:Accession: T01723
A:Status: translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-431 <SCH>
A:Cross-references: EMBL:AF007269; NID:g2191136; PID:g2191143
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 40/2; 94/3; 161/3; 224/2; 255/1; 271/1; 303/1; 339/2
A:Note: A_IG002N01.9

Query Match          94.6%; Score 35; DB 2; Length 431;
Best Local Similarity 85.7%; Pred. No. 6.4;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRXXHC 7
    |||||
Db 383 KKRIEHC 389

RESULT 6
T04537
hypothetical protein F28J12.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 14-May-1999
C:Accession: T04537
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
Submitted to the Protein Sequence Database, February 1998
A:Reference number: Z15377
A:Accession: T04537
A:Molecule type: DNA
A:Residues: 1-633 <BEV>
A:Cross-references: EMBL:AL021710
A:Experimental source: cultivar Columbia; BAC clone F28J12
C:Genetics:
A:Map position: 4
A:Introns: 281/3; 303/3; 442/1; 614/3
A:Note: F28J12.70

Query Match          94.6%; Score 35; DB 2; Length 633;
Best Local Similarity 85.7%; Pred. No. 9;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRXXHC 7
    |||||
Db 591 KKRIMHC 597

RESULT 7
I49668
binding protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999
C:Accession: I49668
R:Hendrickson, B.A.; Zhang, W.; Craig, R.J.; Jin, Y.
Gene 134, 271-275, 1993
A:Title: Structural organization of the genes encoding human and murine FK506-binding
A:Reference number: I49668; MUID:94085790; PMID:7505249
A:Accession: I49668
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-140 <RES>
A:Cross-references: GB:M77831; NID:9433782; PIDN:AAA37631.1; PID:g433783
C:Genetics:
A:Gene: Fkbp13

```

C:Accession: AE2064
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2064
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073766.1; PID:917131158; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2067

Query Match
Best Local Similarity 71.48; Score 33; DB 2; Length 189;
Pred. No. 8.4;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKR1XHC 7
|||: ||
Db 171 KKR1XHC 177

RESULT 13
transposase all1972 [imported] - Nostoc sp. (strain PCC 7120)
AF2052
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AF2052
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2052
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073766.1; PID:917131158; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1972

Query Match
Best Local Similarity 71.48; Score 33; DB 2; Length 189;
Pred. No. 8.4;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKR1XHC 7
|||: ||
Db 171 KKR1XHC 177

RESULT 14
transposase all2773 [imported] - Nostoc sp. (strain PCC 7120)
AF2152
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AF2152
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073766.1; PID:917131158; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2773

A:Gene: all2773

Query Match
Best Local Similarity 71.48; Score 33; DB 2; Length 189;
Pred. No. 8.4;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKR1XHC 7
|||: ||
Db 171 KKR1XHC 177

RESULT 15
transposase all0552 [imported] - Nostoc sp. (strain PCC 7120)
AG1875
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG1875
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1875
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073766.1; PID:917129897; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0552

Query Match
Best Local Similarity 71.48; Score 33; DB 2; Length 189;
Pred. No. 8.4;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKR1XHC 7
|||: ||
Db 171 KKR1XHC 177

Search completed: July 29, 2003, 09:56:56
Job time : 15.1333 secs

A:introns: 55/3, 93/2, 109/1, 121/1
 C:Superfamily: BKBp-type peptidylprolyl isomerase; BKBp-type peptidylprolyl isomerase hc
 F:47-94/Domain: BKBp-type peptidylprolyl isomerase homology <PPI>

Query Match 91.9%; Score 34; DB 2; Length 140;
 Best Local Similarity 71.4%; Pred. No. 4;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKRIXHC 7
 |||: ||
 DB 34 KKRVDHC 40

RESULT 8

FK506/rapamycin-binding protein FKBP13 precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-Aug-1998
 C:Accession: J01365; A39602
 R:Diella, A.G.; Hawkins, A.; Craig, R.J.; Schreiber, S.L.; Griffin, C.A.
 Biochem. Biophys. Res. Commun. 189, 819-823, 1992
 A:Title: Chromosomal band assignments of the genes encoding human FKBP12 and FKBP13.
 A:Reference number: J01365; MUID:93112052; PMID:1281998
 A:Accession: J01365

A:Molecule type: DNA
 A:Residues: 1-142 <DIL>
 R:Jin, Y.J.; Alberts, M.W.; Lane, W.S.; Bierer, B.E.; Schreiber, S.L.; Burakoff, S.J.
 Proc. Natl. Acad. Sci. U.S.A. 88, 6677-6681, 1991
 A:Title: Molecular cloning of a membrane-associated human FK506- and rapamycin-binding F
 A:Reference number: A39602; MUID:93119747; PMID:1713687
 A:Accession: A39602
 A:Molecule type: mRNA
 A:Residues: 1-20, 'S':23-142 <JIN>
 A:Cross-references: GB:M65128
 C:Genetics:

A:Gene: GDB:FKBP2
 A:Cross-references: GDB:133728; OMIM:186946
 A:Map position: 11q13.1-11q13.3
 A:Introns: 57/3, 95/2, 111/1, 123/1
 C:Superfamily: BKBp-type peptidylprolyl isomerase; BKBp-type peptidylprolyl isomerase hc
 C:Keywords: immunoregulation
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-142/Product: FK506/rapamycin-binding protein FKBP13 #status predicted <MAT>
 F:49-96/Domain: BKBp-type peptidylprolyl isomerase homology <PPI>

Query Match 91.9%; Score 34; DB 2; Length 142;
 Best Local Similarity 71.4%; Pred. No. 4;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKRIXHC 7
 |||: ||
 DB 36 KKRVDHC 42

RESULT 9

SE0490
 hypothetical protein YER032w - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C>Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 23-Mar-2001
 C:Accession: S50490
 R:Dieterich, F.S.
 Submitted to the EMBL Data Library, December 1994
 A:Description: The sequence of S. cerevisiae cosmid 9537, 9581, 9495, 9867, and lambda
 A:Reference number: S50433
 A:Accession: S50490
 A:Molecule type: DNA
 A:Residues: 1-925 <DIE>
 A:Cross-references: EMBL:U18778; NID:9603592; PIDN:AAB64565.1; PID:9603624; MIPS:YER032w
 C:Genetics:
 A:Gene: SGD:FIR1
 A:Cross-references: SGD:S0000834; MIPS:YER032w
 A:Map position: 5R

Query Match 91.9%; Score 34; DB 2; Length 925;
 Best Local Similarity 71.4%; Pred. No. 21;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKRIXHC 7
 |||: ||
 DB 841 KKRISHC 847

RESULT 10

A38117
 hypothetical protein 1 - Anabaena sp. insertion sequence IS895
 C:Species: Anabaena sp.
 C>Date: 24-Jul-1992 #sequence_revision 24-Jul-1992 #text_change 15-Oct-1999
 C:Accession: A38117
 R:Alam, J.; Vira, J.M.; Gal, Y.; Martin, J.A.; Weislo, L.J.; Curtis, S.E.
 J. Bacteriol. 173, 5778-5783, 1991
 A:Title: Characterization of the IS895 family of insertion sequences from the cyanoba
 A:Reference number: A38117; MUID:93358370; PMID:1653219
 A:Accession: A38117
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-189 <ALA>

A:Cross-references: GB:M67475; NID:9142026; PIDN:AAA98138.1; PID:9142027
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Mobile element: insertion sequence IS895

Query Match 89.2%; Score 33; DB 2; Length 189;
 Best Local Similarity 71.4%; Pred. No. 8.4;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKRIXHC 7
 |||: ||
 DB 171 KKRKHC 177

RESULT 11

AD2414
 transposase all4868 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AD2414
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matsumoto, A.; Iriku
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD2414
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-189 <KUP>
 A:Cross-references: GB:BA000019; PIDN:BA076567.1; PID:917134005; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all4868

Query Match 89.2%; Score 33; DB 2; Length 189;
 Best Local Similarity 71.4%; Pred. No. 8.4;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKRIXHC 7
 |||: ||
 DB 171 KKRKHC 177

RESULT 12

AE2064
 transposase all2067 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:52:17 ; Search time 7 Seconds
(without alignments)
47.027 Million cell updates/sec

Title: US-09-606-129A-16

Perfect score: 37

Sequence: 1 KKRIHXC 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	35	94.6	295	1 BIEA_RAT	P46844 rattus norv
2	35	94.6	296	1 BIEA_HUMAN	P53004 homo sapien
3	34	91.9	140	1 FRB2_MOUSE	P43678 mus musculu
4	34	91.9	142	1 FRB2_HUMAN	P26885 homo sapien
5	34	91.9	925	1 PIP1_YEAST	P40020 saccharomyc
6	33	89.2	323	1 CYCH_XENLA	P51947 xenopus lae
7	33	89.2	473	1 OASL_MOUSE	O92f22 mus musculu
8	30	81.1	38	1 RL36_CHLTE	O8kaJ4 chlorobium
9	30	81.1	38	1 RL36_THEMA	O9xli6 thermotoga
10	30	81.1	108	1 YZ09_METJA	O60271 methanococc
11	30	81.1	528	1 DYR2_HUMAN	O92630 homo sapien
12	30	81.1	571	1 AMPL_LVCES	Q10712 lycopersico
13	30	81.1	573	1 AMPL_SOLTU	P31427 solanum tub
14	30	81.1	861	1 UL52_HSV7J	P52468 human herpe
15	30	81.1	4128	1 PRKD_MOUSE	P97313 mus musculu
16	29	78.4	142	1 SECB_BUCAI	P57161 buchnera ap
17	29	78.4	451	1 SUN_HAEIN	P44788 haemophilus
18	29	78.4	477	1 URT1_DESRO	P98119 desmodus ro
19	29	78.4	477	1 URT2_DESRO	P15638 desmodus ro
20	29	78.4	621	1 GIDA_CHLTE	O8ka85 chlorobium
21	29	78.4	632	1 AFUB_HAEIN	O57341 haemophilus
22	29	78.4	899	1 R24L_ARATH	O9c646 arabidopsis
23	29	78.4	1220	1 IF2P_HUMAN	O60841 homo sapien
24	29	78.4	3148	1 HD_FUGRU	P51112 fuqu rubrip
25	28	75.7	324	1 D3H1_DROME	O9v8m5 drosophila
26	28	75.7	359	1 ODPB_RAT	P49432 rattus norv
27	28	75.7	372	1 YEPI_ECOLI	P37751 escherichia
28	28	75.7	586	1 HO_YEAST	P09932 saccharomyc
29	28	75.7	835	1 SYFB_COREF	O8ftp0 corynebacte
30	28	75.7	835	1 SYFB_CORGL	O8nqn6 corynebacte
31	28	75.7	875	1 HELX_SULSO	P95949 sulfolobus
32	28	75.7	900	1 MUTS_TREPA	O83348 treponema p
33	28	75.7	3110	1 HD_RAT	P51111 rattus norv

RESULT 1

ID	BIEA_RAT	STANDARD;	PRT;	295 AA.
AC	P46844;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Biliverdin reductase A precursor (EC 1.3.1.24) (Biliverdin-IX alpha-reductase).			
DE	reductase).			
GN	BLVRA OR BLVR.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	TISSUE=Kidney;			
RX	MEDLINE=92156147; PubMed=1371282;			
RA	Fakhrai H., Maines M.D.;			
RT	"Expression and characterization of a cDNA for rat kidney biliverdin reductase. Evidence suggesting the liver and kidney enzymes are the same transcript product."			
RT	J. Biol. Chem. 267:4023-4029(1992).			
RN	[2]			
RP	MUTAGENESIS.			
RX	MEDLINE=94291657; PubMed=8020496;			
RA	McCoubrey W.K. Jr., Maines M.D.;			
RT	"Site-directed mutagenesis of cysteine residues in biliverdin reductase. Roles in substrate and cofactor binding."			
RL	Eur. J. Biochem. 222:597-603(1994).			
CC	!- FUNCTION: CONVERTS BILIVERDIN TO BILIRUBIN; DISPLAYS TWO DISTINCT PH OPTIMA USING A DIFFERENT COFACTOR AT EACH PH: NADH AT THE LOWER PH 6.7-6.9 RANGE AND NADPH AT PH 8.5-8.7. NADPH, HOWEVER, IS THE PROBABLE COFACTOR IN BIOLOGICAL SYSTEMS.			
CC	!- CATALYTIC ACTIVITY: Bilirubin + NAD(P)(+) -> biliverdin + NAD(P)H.			
CC	!- COFACTOR: BINDS ONE ZINC ION.			
CC	!- PATHWAY: Heme metabolism; first step.			
CC	!- SUBUNIT: Monomer (By similarity).			
CC	!- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	!- SIMILARITY: TO E.COLI YHHX.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M81681; AAA40830.1; -			
DR	PIR; A42268; A42268.			
DR	PDB; 1GCU; 28-JAN-03.			
DR	PDB; 1LC0; 17-JUL-02.			
DR	PDB; 1LC3; 17-JUL-02.			
DR	InterPro; IPR000683; GFO_IDH_MoCA.			
DR	Pfam; PF01408; GFO_IDH_MoCA; 1.			

ALIGNMENTS

34	28	75.7	3119	1 HD_MOUSE	P42859 mus musculu
35	28	75.7	3144	1 HD_HUMAN	P42858 homo sapien
36	27	73.0	89	1 RL34_PYRFU	O8u213 pyrococcus
37	27	73.0	91	1 RL34_PYRAB	O9uz17 pyrococcus
38	27	73.0	91	1 RL34_PYRHO	O74006 pyrococcus
39	27	73.0	213	1 NO27_SOYBN	P08864 glycine max
40	27	73.0	214	1 SODM_PASMU	O9cpn6 pasteurella
41	27	73.0	215	1 NO22_SOYBN	P08961 glycine max
42	27	73.0	308	1 MENA_HAEIN	P44739 haemophilus
43	27	73.0	331	1 YB52_SYNV3	P74221 synechocyst
44	27	73.0	374	1 TF3A_SCHPO	O9ut15 schizosach
45	27	73.0	457	1 DNAA_HELPJ	O9zj96 helicobacte


```

Db          275 KKRILHC 281

RESULT 3
FKB2_MOUSE
ID_FKB2_MOUSE STANDARD; PRT; 140 AA.
AC P45878;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE FK506-binding protein 2 precursor (EC 5.2.1.8) (peptidyl-prolyl cis-
DE trans isomerase) (PPIase) (Rotamase) (13 kDa FKBP) (FKBP-13).
GN FKBp2 OR FKBp13.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ; TISSUE=Liver;
RX MEDLINE=94085790; PubMed=7505249;
RA Hendrickson B.A., Zhang W., Craig R.J., Jin Y.J., Bierer R.E.,
RA Burakoff S.J., Dilella A.G.;
RT "Structural organization of the genes encoding human and murine
RT FK506-binding protein (FKBP) 13 and comparison to FKBp1.";
RL Gene 134:271-275(1993).
CC -!- FUNCTION: PPIases accelerate the folding of proteins. It catalyzes
CC the cis-trans isomerization of proline imidic peptide bonds in
CC oligopeptides.
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC -!- SUBUNIT: Interacts with the C-terminal domain of 4.1g.
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN. MEMBRANE
CC ASSOCIATED (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M77831; AAA37631.1; -
CC PIR; I49668; I49668.
CC HSP; P20081; 1YAT.
CC MGP; MGI:95542; Fkbp2.
CC InterPro; IPR001179; FKBP_PPIase.
CC Pfam; PF00254; FKBP; 1.
CC PROSITE; PS00453; FKBP_PPIASE.1; 1.
CC PROSITE; PS00454; FKBP_PPIASE.2; 1.
CC PROSITE; PS50059; FKBP_PPIASE.3; 1.
KW Isomerase; Rotamase; Signal; Endoplasmic reticulum.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 140 FK506-BINDING PROTEIN 2.
FT SITE 137 140 PREVENT SECRETION FROM ER (POTENTIAL).
FT TYPE 140
SQ SEQUENCE 140 AA; 15344 MW; F4E7FCC7766A0416 CRC64;

Query Match 91.9%; Score 34; DB 1; Length 140;
Best Local Similarity 71.4%; Pred. No. 1;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KKRXXHC 7
III: ||
Db 34 KKRVDHC 40

RESULT 4
FKB2_HUMAN
ID_FKB2_HUMAN STANDARD; PRT; 142 AA.
AC P26885; Q9BTS7;

```


CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).

DR EMBL: U20505; AAA62236.1; -
 DR HSP: P51946; LJM
 DR InterPro: IPR005258; Cell.
 DR InterPro: IPR006670; Cyclin.
 DR InterPro: IPR006671; Cyclin_N.
 DR Pfam: PF00134; cyclin; 1.
 DR SMART: SM00385; CYCLIN; 1.
 DR TIGRfams: TIGR00569; cell; 1.
 DR PROSITE: PS00292; CYCLINS; FALSE_NEG.
 KW Transcription regulation; Cell cycle; Nuclear protein; Cyclin;
 FT Phosphorylation.
 FT CONFLICT 266
 SQ SEQUENCE 323 AA; 37600 MW; 14BCDCA00843DC8 CRC64;

Query Match 89.2%; Score 33; DB 1; Length 323;
 Best Local Similarity 71.4%; Pred. No. 4;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIKHC 7
 III: II
 DB 274 KKRLDHC 280

RESULT 7
 ID OASL_MOUSE STANDARD; PRT: 473 AA.
 AC Q922F2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 54 kDa 2'-5'-oligoadenylate synthetase like protein (EC 2.7.7.-) (p54
 DE OASL) (P54OASL) (M1204).
 GN OASL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Spleen;
 RX MEDLINE=99323969; PubMed=10393668;
 RA Tienfenthaler M., Marksteiner R., Neyer S., Koch F., Hofer S.,
 RA Schuler G., Nussenzweig M., Schneider R., Heufler C.;
 RT "M1204, a novel 2'-5'-oligoadenylate synthetase with a ubiquitin-like
 RT extension, is induced during maturation of murine dendritic cells.";
 RL J. Immunol. 163:760-765(1999).
 CC -!- FUNCTION: MAY PLAY A ROLE IN MEDIATING RESISTANCE TO VIRUS
 CC INFECTION, CONTROL OF CELL GROWTH, DIFFERENTIATION, AND APOPTOSIS.
 CC -!- CATALYTIC ACTIVITY: Binds double-stranded RNA and polymerizes ATP
 CC into ppp(A2'p5'A)n oligomers, which activate the latent RNASE L
 CC that, when activated, cleaves single-stranded RNAs.
 CC -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN SPLEEN DENDRITIC CELLS,
 CC WHEREAS, IN BONE MARROW-DERIVED DENDRITIC CELLS, THE AMOUNT
 CC INCREASES DURING THE MATURATION PROCESS. EXPRESSED IN MANY ORGANS,
 CC THE HIGHEST LEVELS BEING IN THYMUS, LUNG, AND BONE MARROW.
 CC -!- SIMILARITY: BELONGS TO THE 2-5A SYNTHETASE FAMILY.
 CC -!- SIMILARITY: Contains 1 ubiquitin-like domain.
 CC -!- CAUTION: THIS MAY NOT BE THE TRUE ORTHOLOG OF HUMAN OASL.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF068835; AAD02818.1; -
 DR MGDI: MGDI344390; Oasl.
 DR GO: GO:0001730; F:2'-5'-oligoadenylate synthetase activity; IDA.

DR InterPro: IPR006117; 25A_SYNTH_2.
 DR InterPro: IPR006116; 25A_synth_UB.
 DR InterPro: IPR001201; PAP_25A_core.
 DR InterPro: IPR000626; Ubiquitin.
 DR Pfam: PF00240; ubiquitin; 1.
 DR PROSITE: PS00832; 25A_SYNTH_1; FALSE_NEG.
 DR PROSITE: PS00833; 25A_SYNTH_2; 1.
 DR PROSITE: PS0152; 25A_SYNTH_3; 1.
 DR PROSITE: PS0053; UBIQUITIN_2; FALSE_NEG.
 KW RNA-binding; Transferase; Nucleotidyltransferase.
 FT DOMAIN 435 473 UBIQUITIN-LIKE
 SQ SEQUENCE 473 AA; 54625 MW; 570E0E08A51C8460 CRC64;

Query Match 89.2%; Score 33; DB 1; Length 473;
 Best Local Similarity 71.4%; Pred. No. 5.9;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKRIKHC 7
 III: II
 DB 109 KKRLIHC 115

RESULT 8
 ID RL36_CHLTFE STANDARD; PRT: 38 AA.
 AC Q8KAJ4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L36.
 GN RPMJ OR CT2166.
 OS Chlorobium tepidum.
 OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
 OC Chlorobium.
 OX NCBI_TaxID=1097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TLS / ATCC 49652 / DSM 12025;
 RX MEDLINE=22103685; PubMed=12093901;
 RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
 RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
 RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
 RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
 RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
 RA Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
 RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
 RT "The complete genome sequence of Chlorobium tepidum TLS, a
 RT photosynthetic, anaerobic, green-sulfur bacterium";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
 CC -!- SIMILARITY: BELONGS TO THE L36P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE012964; AAM73382.1; -
 DR TIGR: CT2166; -
 DR HAMAP: MF_00251; -; 1.
 DR InterPro: IPR000473; Ribosomal_L36.
 DR Pfam: PF00444; Ribosomal_L36; 1.
 DR ProDom: PD002101; Ribosomal_L36; 1.
 DR TIGRfams: TIGR01022; rpmJ_bact; 1.
 DR PROSITE: PS00828; RIBOSOMAL_L36; FALSE_NEG.
 KW Ribosomal protein; Complete proteome.
 SQ SEQUENCE 38 AA; 4586 MW; 40EAF708FE4EE21 CRC64;

Query Match 81.1%; Score 30; DB 1; Length 38;
 Best Local Similarity 71.4%; Pred. No. 2.1;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Oy 1 KKRXXHC 7
Db 8 KKRCEHC 14

RESULT 9
RL36_THEME
ID RL36_THEME STANDARD; PRT; 38 AA.
AC Q9XLI6; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L36.
GN RPMJ OR TM1476
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
SEQUENCE FROM N.A.
STRAIN=MSB8 / DSM 3109;
MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Hatt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
"Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RC Nature 399:323-329(1999).
CC -!- SIMILARITY: BELONGS TO THE L36P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001798; AAD36568.1; -
CC PIR; E72247; E72247.
CC HSP; P80256; 1DFE.
CC TIGR; TM1476; -.
CC HAMAP; MF_00251; -.
CC InterPro; IPR000473; Ribosomal_L36.
CC Pfam; PF00444; Ribosomal_L36; 1.
CC ProDom; PD002101; Ribosomal_L36; 1.
CC TIGRFAMs; TIGR01022; rpmj_bact; 1.
CC PROSITE; PS00828; RIBOSOMAL_L36; 1.
CC Ribosomal protein; Complete proteome.
CC SEQUENCE 38 AA; 4564 MW; 263FDD759798800B CRC64;

Query Match 81.1%; Score 30; DB 1; Length 38;
Best Local Similarity 71.4%; Pred. No. 2.1;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KKRXXHC 7
Db 8 KKRCEHC 14

RESULT 10
YZ09_METJA
ID YZ09_METJA STANDARD; PRT; 108 AA.
AC Q60271;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJEC109.
GN MJEC109.
OS Methanococcus jannaschii.

Archaea; Euryarchaeota; Methanococci; Methanococcales;
Methanocaldococaceae; Methanocaldococcus.
NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A.
STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
Overbeek R., Kirkness E.F., Weinstock K.F., Merrick J.M., Glodek A.,
Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
"Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RC Science 273:1058-1073(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L77118; AAC37082.1; -
CC PIR; A64511; A64511.
CC TIGR; MJEC109; -.
CC Hypothetical protein; Complete proteome.
CC SEQUENCE 108 AA; 13308 MW; 482EE7DFA65A277D CRC64;

Query Match 81.1%; Score 30; DB 1; Length 108;
Best Local Similarity 83.3%; Pred. No. 6;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 KKRXXHC 7
Db 80 KKRXXHC 85

RESULT 11
DYR2_HUMAN
ID DYR2_HUMAN STANDARD; PRT; 528 AA.
AC Q92630;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dual-specificity tyrosine-phosphorylation regulated kinase 2
DE (EC 2.7.1.1).
GN DYRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98421512; PubMed=9748265;
RA Becker W., Weber Y., Wetzel K., Eirmbter K., Tejedor F.J.,
RA Joost H.-G.;
RT "Sequence characteristics, subcellular localization, and substrate
RT specificity of DYRK-related kinases, a novel family of dual
RT specificity protein kinases."
RL J. Biol. Chem. 273:25893-25902(1998).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

```

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 320-528 FROM N.A.
 RC TISSUE=Placenta;
 RA Becker W., Joost H.-G.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: IN VITRO; CAN PHOSPHORYLATE HISTONES H3 AND H2B ON SER
 CC AND THR RESIDUES. MAY BE INVOLVED IN THE REGULATION OF CELLULAR
 CC GROWTH AND/OR DEVELOPMENT.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MNE/DYRK SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y13493; CAA73885.1; -;
 DR EMBL; BC005809; AAH05809.1; -;
 DR EMBL; Y09216; CAA70418.1; -;
 DR HSSP; Q00534; IRI8.
 DR Genew; HGNC:3093; DYRK2.
 DR MIM; 603496; -;
 DR GO; GO:0005737; C:cytoplasm; TAS.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_Kinase; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW Transferase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
 KW ATP-binding; Phosphorylation.
 FT DOMAIN 149 462 PROTEIN KINASE.
 FT NP_BIND 155 163 ATP (BY SIMILARITY).
 FT BINDING 178 178 ATP (BY SIMILARITY).
 FT ACT_SITE 275 275 BY SIMILARITY.
 SQ SEQUENCE 528 AA: 59714 MW: AF2C6822ED952D7 CRC64;
 Query Match 81.1%; Score 30; DB 1; Length 528;
 Best Local Similarity 71.4%; Pred. NO. 30;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KKRXXHC 7
 Db 268 KNRIHC 274
 RESULT 12
 AMPL_LYCES

ID AMPL_LYCES STANDARD; PRT; 571 AA.
 AC Q10712; Q3S9A3;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aminopeptidase 1, chloroplast precursor (EC 3.4.11.1) (Leucine
 DE aminopeptidase) (LAP) (Leucyl aminopeptidase) (Proline aminopeptidase)
 DE (EC 3.4.11.5) (Prolyl-aminopeptidase) (DR57).
 GN LAPA1 OR LAP OR LAP2
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. PETO 238R; TISSUE=Leaf;
 RX MEDLINE=96421572; PubMed=8824220;
 RA Gu Y.O., Chao W.S., Walling L.L.;
 RT "Localization and post-translational processing of the wound-induced
 RT leucine aminopeptidase proteins of tomato";
 RL J. Biol. Chem. 271:25880-25887(1996).
 RN [2]
 RP SEQUENCE OF 49-571 FROM N.A.
 RC STRAIN=CV. VF36; TISSUE=Pistil;
 RX MEDLINE=95375233; PubMed=7647301;
 RA Milligan S.B., Gasser C.S.;
 RT "Nature and regulation of pistil-expressed genes in tomato";
 RL Plant Mol. Biol. 28:691-711(1995).
 RN [3]
 RP SEQUENCE OF 103-571 FROM N.A.
 RC STRAIN=CV. PETO 238R; TISSUE=Leaf;
 RX MEDLINE=94052201; PubMed=8234334;
 RA Pautot V., Holzer F.M., Reisch B., Walling L.L.;
 RT "Leucine aminopeptidase: an inducible component of the defense
 RT response in Lycopersicon esculentum (tomato).";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:9906-9910(1993).
 CC -!- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
 CC TURNOVER OF INTRACELLULAR PROTEINS.
 CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-l-
 CC xbb-, in which Xaa is preferably Leu, but may be other amino acids
 CC including Pro although not Arg or Lys, and Xbb may be Pro.
 CC -!- CATALYTIC ACTIVITY: Release of a N-terminal proline from a
 CC peptide.
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Chloroplast.
 CC -!- INDUCTION: By wounding.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U50151; AAC49456.1; -;
 DR EMBL; U50152; AAC49457.1; -;
 DR EMBL; U20593; AAA80498.1; -;
 DR PIR; T07849; T07849.
 DR HSSP; P00727; LBPN.
 DR MEROPS; M17.002; -;
 DR InterPro; IPR000819; Peptidase_M17.
 DR Pfam; PF00883; Peptidase_M17; 1.
 DR Pfam; PF02789; Peptidase_M17_N; 1.
 DR PRINTS; PR00481; LAMNOPPTDASE.
 DR PROSITE; PS00631; CYTOSOLAP; 1.
 KW Transit peptide; Chloroplast; Aminopeptidase; Hydrolase; Zinc.
 FT TRANSIT 1 53 CHLOROPLAST (POTENTIAL).
 FT CHAIN 54 571 AMINOPEPTIDASE 1.
 FT DOMAIN 169 174 POLY-ALA.
 FT METAL 342 342 ZINC 2 (BY SIMILARITY).

FT METAL 347 347 ZINC 1 AND 2 (BY SIMILARITY).
 FT METAL 367 367 ZINC 2 (BY SIMILARITY).
 FT METAL 427 427 ZINC 1 (BY SIMILARITY).
 FT METAL 429 429 ZINC 1 AND 2 (BY SIMILARITY).
 FT ACT_SITE 354 354 POTENTIAL.
 FT ACT_SITE 431 431 R -> G.
 FT VARIANT 358 358 P -> N (IN REF. 2).
 FT CONFLICT 271 271 T -> S (IN REF. 3).
 FT CONFLICT 315 315 T -> L (IN CLONE PBLAP2).
 FT CONFLICT 515 515 T -> L (IN CLONE PBLAP2).
 SQ SEQUENCE 571 AA; 60279 MW; C7A224837E73939D CRC64;

Query Match 81.1%; Score 30; DB 1; Length 571;
 Best Local Similarity 83.3%; Pred. No. 32;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KR1XHC 7
 Db 48 KR1VHC 53

RESULT 13
 AMPL_SOLUTU STANDARD; PRT; 573 AA.
 ID AMPL_SOLUTU STANDARD; PRT; 573 AA.
 AC P31427;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Aminopectidase, chloroplast precursor (EC 3.4.11.1) (Leucine
 DE aminopectidase) (LAP) (leucyl aminopectidase) (Proline aminopectidase)
 DE (EC 3.4.11.5) (Prolyl aminopectidase).
 GN LAP.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
 OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Desiree;
 RX MEDLINE=94339796; PubMed=7765119;
 RA Herbers K., Prat S., Willmitzer L.;
 RT "Functional analysis of a leucine aminopectidase from Solanum
 RL tuberosum L.";
 RL Planta 194:230-240(1994).
 RN [2]
 RP SEQUENCE OF 19-573 FROM N.A.
 RC STRAIN=cv. Desiree; TISSUE=Leaf;
 RX MEDLINE=93005746; PubMed=1392612;
 RA Hildmann T., Ebner M., Pena-Cortes H., Sanchez-Serrano J.J.,
 RA Willmitzer L., Prat S.;
 RT "General roles of abscisic and jasmonic acids in gene activation as a
 RT result of mechanical wounding.";
 RL Plant Cell 4:1157-1170(1992).
 CC -!- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
 CC TURNOVER OF INTRACELLULAR PROTEINS.
 CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-[
 CC Xbb-, in which Xaa is preferably Leu, but may be other amino acids
 CC including Pro although not Arg or Lys, and Xbb may be Pro.
 CC -!- CATALYTIC ACTIVITY: Release of a N-terminal proline from a
 CC peptide.
 CC -!- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -!- SUBUNIT: Homohexamer (Probable).
 CC -!- SUBCELLULAR LOCATION: Chloroplast (By similarity).
 CC -!- TISSUE SPECIFICITY: IN TUBERS AND FLORAL BUDS OF UNTREATED PLANTS.
 CC AFTER ABA TREATMENT OR MECHANICAL WOUNDING IS MOSTLY ACCUMULATED
 CC IN LEAVES, TO A LESSER EXTENT IN STEMS, BUT NOT IN ROOTS.
 CC -!- INDUCTION: BY ABSICISIC ACID (ABA), JASMONIC ACID (JA) AND
 CC WOUNDING.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X77015; CAA54314.1; -;
 CC DR EMBL; X67845; CAA48038.1; -;
 CC DR PIR; S41376; S41376.
 CC DR HSP; P00727; ILAM.
 CC DR MEROPS; M17.002; -;
 CC DR InterPro; IPR00819; Peptidase_M17.
 CC DR Pfam; PF00883; Peptidase_M17; 1.
 CC DR Pfam; PF02789; Peptidase_M17.N; 1.
 CC DR PRINTS; PF00481; LAMNOPPTDASE.
 CC DR PROSITE; PS00631; CYTOSOLAP; 1.
 CC DR Transil peptide; Chloroplast; Aminopectidase; Hydrolase; Zinc.
 CC FT TRANSIT 1 53 CHLOROPLAST (POTENTIAL).
 CC FT CHAIN 54 573 AMINOPEPTIDASE.
 CC FT DOMAIN 169 174 POLY-ALA.
 CC FT METAL 342 342 ZINC 2 (BY SIMILARITY).
 CC FT METAL 347 347 ZINC 1 AND 2 (BY SIMILARITY).
 CC FT METAL 367 367 ZINC 2 (BY SIMILARITY).
 CC FT METAL 427 427 ZINC 1 (BY SIMILARITY).
 CC FT METAL 429 429 ZINC 1 AND 2 (BY SIMILARITY).
 CC FT ACT_SITE 354 354 POTENTIAL.
 CC FT ACT_SITE 431 431 POTENTIAL.
 SQ SEQUENCE 573 AA; 60122 MW; 3152145A4A7FB291 CRC64;

Query Match 81.1%; Score 30; DB 1; Length 573;
 Best Local Similarity 83.3%; Pred. No. 32;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KR1XHC 7
 Db 48 KR1VHC 53

RESULT 14
 UL52_HSV7J STANDARD; PRT; 861 AA.
 ID UL52_HSV7J
 AC P52468;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Helicase/primase complex protein.
 GN U43.
 OS Human herpesvirus (type 7 / strain J1) (HHV7).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 OX NCBI_TaxID=57278;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nicholas J.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: INVOLVED IN DNA REPLICATION.
 CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL52,
 CC EBV-1 7, EBV BSFL1, HVS-1 56, HCMV UL70 AND VZV 6.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U43400; AAC54705.1; -;
 CC DR PIR; T41945; T41945.
 CC DR InterPro; IPR004340; UL52_UL70.
 CC DR Pfam; PF03121; UL52_UL70; 1.
 CC DR DNA replication.
 SQ SEQUENCE 861 AA; 101073 MW; 706B74EB4A007EE3 CRC64;

Query Match 81.1%; Score 30; DB 1; Length 861;
 Best Local Similarity 71.4%; Pred. No. 49;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KKRIHC 7
 Db 763 KNIHC 769

RESULT 15
 ID PRKD_MOUSE STANDARD; PRT: 4128 AA.
 AC 97313; 088187; P97928; Q92341;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA-dependent protein kinase catalytic subunit (EC 2.7.1.37) (DNA-
 DE PKCS) (P4560).
 GN PRKDC OR XRCC7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

[1]
 RN SEQUENCE FROM N.A., AND VARIANT SCID.
 RP STRAIN=C.B17; TISSUE=Fibroblast, and Leukocyte;
 RC MEDLINE=97225971; PubMed=9122213;
 RA Araki R., Fujimori A., Hanatani K., Mita K., Saito T., Mori M.,
 RA Fukumura R., Morimyo M., Muto M., Itoh M., Tatsumi K., Abe M.;
 RT "Nonsense mutation at Tyr-4046 in the DNA-dependent protein kinase
 RT catalytic subunit of severe combined immune deficiency mice.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:2438-2443(1997).
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98250757; PubMed=9582343;
 RA Fukumura R., Araki R., Fujimori A., Mori M., Saito T., Watanabe F.,
 RA Sarashi M., Itsukaichi H., Eguchi-Kasai K., Sato K., Tatsumi K.,
 RA Abe M.;
 RT "Murine cell line SX9 bearing a mutation in the DNA-PKcs gene exhibits
 RT aberrant V(D)J recombination not only in the coding joint but also in
 RT the signal joint.";
 RL J. Biol. Chem. 273:13058-13064(1998).
 [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=129/SVJ;
 RC MEDLINE=97480731; PubMed=9339376;
 RA Fujimori A., Araki R., Fukumura R., Saito T., Mori M., Mita K.,
 RA Tatsumi K., Abe M.;
 RT "The murine DNA-PKcs gene consists of 86 exons dispersed in more than
 RT 250 kb.";
 RL Genomics 45:194-199(1997).
 [4]
 RN SEQUENCE OF 1-26 FROM N.A.
 RP MEDLINE=98384545; PubMed=9716665;
 RA Saito T., Matsuda Y., Ishii H., Watanabe F., Mori M., Hayashi A.,
 RA Araki R., Fujimori A., Fukumura R., Morimyo M., Tatsumi K., Hori T.,
 RA Abe M.;
 RT "Mouse Cdc21 only 0.5 kb upstream from DNA-PKcs in a head-to-head.
 RT organization: an implication of co-evolution of ATM family members and
 RT cell cycle regulating genes.";
 RL Mamm. Genome 9:769-772(1998).
 [5]
 RN SEQUENCE OF 3615-4128 FROM N.A.
 RP STRAIN=BALB/c; TISSUE=Leukocyte;
 RC MEDLINE=97045026; PubMed=8881030;
 RA Hanatani K., Matsuda Y., Araki R., Itoh M., Abe M.;
 RA "Cloning and chromosomal mapping of the mouse DNA-dependent protein
 RT kinase gene.";
 RL Immunogenetics 45:1-5(1996).
 [6]
 RN SEQUENCE OF 3680-4128 FROM N.A.
 RP MEDLINE=96413638; PubMed=8816792;

RA Blunt T., Gell D., Fox M., Taccioli G.E., Lehmann A.R., Jackson S.P.,
 RA Jeggo P.A.;
 RT "Identification of a nonsense mutation in the carboxyl-terminal region
 RT of DNA-dependent protein kinase catalytic subunit in the scid mouse.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:10285-10290(1996).
 [7]
 RN SEQUENCE OF 3839-4128 FROM N.A.
 RP STRAIN=C.B17;
 RC MEDLINE=96413301; PubMed=8816463;
 RA Danska J.S., Holland D.P., Mariathasan S., Williams K.M., Guidos C.J.;
 RT "Biochemical and genetic defects in the DNA-dependent protein kinase
 RT in murine scid lymphocytes.";
 RL Mol. Cell. Biol. 16:5507-5517(1996).
 CC -!- FUNCTION: SER/THR KINASE INVOLVED IN DNA DOUBLE-STRANDED BREAK
 CC REPAIR. V(D)J RECOMBINATION AND MODULATION OF TRANSCRIPTION. MUST
 CC BE BOUND TO DNA TO EXPRESS ITS CATALYTIC PROPERTIES.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein -> ADP + a phosphoprotein.
 CC -!- SUBUNIT: FORMS A COMPLEX WITH THE HETERODIMER KU70/KU80
 CC AUTOANTIGEN. INTERACTS WITH DNA-PKCS INTERACTING PROTEIN (KIP)
 CC WITH THE REGION UPSTREAM THE KINASE DOMAIN.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- DISEASE: DEFECTS IN PRKDC ARE THE CAUSE OF SEVERE COMBINED IMMUNE
 CC DEFICIENCY (SCID) WHICH IS CHARACTERIZED BY A LACK OF MATURE
 CC FUNCTIONAL LYMPHOCYTES AND A HIGH SUSCEPTIBILITY TO LETHAL
 CC OPPORTUNISTIC INFECTIONS IF NOT CHRONICALLY TREATED WITH
 CC ANTIBIOTICS. THE LACK OF B- AND T-CELL IMMUNITY RESEMBLES SEVERE
 CC COMBINED IMMUNODEFICIENCY (SCID) SYNDROME IN HUMAN INFANTS.
 CC -!- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; D87521; BAA19566.1; -;
 DR EMBL; AB007544; BAA28873.1; -;
 DR EMBL; AB030754; BAB91149.1; -;
 DR EMBL; AB011543; BAA28875.1; -;
 DR EMBL; AB000629; BAA34640.1; -;
 DR EMBL; D83786; BAA12115.1; -;
 DR EMBL; U78157; AAB36939.1; -;
 DR EMBL; U78158; AAB36940.1; -;
 DR PIR; JC6306; JC6306;
 DR MGI; MGI:104779; Prkdc.
 DR InterPro; IPR003151; FAT.
 DR InterPro; IPR003152; FATC.
 DR InterPro; IPR000403; PI3_P14_Kinase.
 DR Pfam; PF02259; FAT; 1.
 DR Pfam; PF02260; FATC; 1.
 DR Pfam; PF00454; PI3_P14_kinase; 1.
 DR SMART; SM00146; PI3Kc; 1.
 DR PROSITE; PS00915; PI3_4_KINASE.1; 1.
 DR PROSITE; PS00916; PI3_4_KINASE.2; FALSE_NEG.
 DR PROSITE; PS02090; PI3_4_KINASE.3; 1.
 KW Transferrase; Serine/threonine-protein kinase; DNA repair;
 KW Nuclear protein; Disease mutation; SCID; Polymorphism;
 KW Phosphorylation.
 FT DOMAIN 3748 4128
 FT MOD_RES 2605
 FT MOD_RES 2608
 FT MOD_RES 2634
 FT MOD_RES 2643
 FT VARIANT 2140 2140
 FT VARIANT 3191 3191
 FT VARIANT 4046 4128
 FT CONFLICT 3844 3844
 FT CONFLICT 4128 AA; 471383 MW; C2D7368D8E50AEE0 CRC64;
 FT SEQUENCE

Query Match 81.1%; Score 30; DB 1; Length 4128;
 Best Local Similarity 57.1%; Pred. No. 2.4e+02;

Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Oy 1 KKRIXHC 7
|:|: ||
Db 1933 KRRLYHC 1939

Search completed: July 29, 2003, 09:55:17
Job time : 9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:52:37 ; Search time 28 Seconds
(without alignments)
64,513 Million cell updates/sec

Title: US-09-606-129A-16
Perfect score: 37
Sequence: 1 KKR1XHC 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	94.6	295	11 Q9CY64	Q9CY64 mus musculu
2	35	94.6	303	11 Q9DD21	Q9DD21 mus musculu
3	35	94.6	401	10 Q9M149	Q9M149 arabidopsis
4	35	94.6	431	10 Q04613	Q04613 arabidopsis
5	35	94.6	599	10 Q9FHV3	Q9FHV3 arabidopsis
6	35	94.6	633	10 Q49511	Q49511 arabidopsis
7	33	89.2	189	16 Q03949	Q03949 anabaena sp
8	33	89.2	508	11 Q9D6S2	Q9D6S2 mus musculu
9	33	89.2	508	11 Q8K234	Q8K234 mus musculu
10	33	89.2	510	10 Q8L5V3	Q8L5V3 arabidopsis
11	33	89.2	510	10 Q8GWA1	Q8GWA1 arabidopsis
12	33	89.2	512	10 Q9LML0	Q9LML0 arabidopsis
13	33	89.2	562	10 Q8H5X6	Q8H5X6 oryza sativ
14	32	86.5	109	5 Q9VZN4	Q9VZN4 drosophila
15	32	86.5	683	6 Q9N091	Q9N091 macaca fasc
16	32	86.5	1231	10 Q9FKE2	Q9FKE2 arabidopsis

17	31	83.8	106	5 Q9XZM5	Q9XZM5 leishmania
18	31	83.8	160	5 Q8WS63	Q8WS63 glossina mo
19	31	83.8	175	16 Q8CVV2	Q8CVV2 escherichia
20	31	83.8	247	16 Q8F298	Q8F298 leptospira
21	31	83.8	279	11 Q9QXV3	Q9QXV3 mus musculu
22	31	83.8	321	10 Q9LZD9	Q9LZD9 arabidopsis
23	31	83.8	457	5 Q93565	Q93565 caenorhabdi
24	31	83.8	712	16 Q8D321	Q8D321 wiggleswort
25	31	83.8	852	5 Q8WSN1	Q8WSN1 drosophila
26	31	83.8	853	5 Q8WR83	Q8WR83 drosophila
27	31	83.8	853	5 Q9V6E4	Q9V6E4 drosophila
28	31	83.8	873	17 Q971E9	Q971E9 sulfolobus
29	29	83.1	95	9 Q38306	Q38306 lactococcus
30	30	81.1	102	10 Q9S738	Q9S738 lycopersico
31	30	81.1	143	2 Q9FC22	Q9FC22 erwinia ste
32	30	81.1	197	16 Q97P95	Q97P95 streptococc
33	30	81.1	197	16 Q8DNQ7	Q8DNQ7 streptococc
34	30	81.1	221	16 Q8E955	Q8E955 shewanella
35	30	81.1	252	10 Q945Y2	Q945Y2 theobroma c
36	30	81.1	267	2 Q9AKS0	Q9AKS0 pseudomonas
37	30	81.1	302	16 Q98H23	Q98H23 rhizobium l
38	30	81.1	352	5 Q9VWP7	Q9VWP7 drosophila
39	30	81.1	388	2 Q8LOM8	Q8LOM8 streptococc
40	30	81.1	411	10 Q9LMM1	Q9LMM1 arabidopsis
41	30	81.1	427	10 Q23705	Q23705 arabidopsis
42	30	81.1	463	16 Q9K7N7	Q9K7N7 bacillus ha
43	30	81.1	465	5 Q9Y029	Q9Y029 drosophila
44	30	81.1	495	10 Q9ST63	Q9ST63 solanum tub
45	30	81.1	499	10 Q941V1	Q941V1 oryza sativ

ALIGNMENTS

RESULT 1

ID	Q9CY64	PRELIMINARY:	PRT:	295 AA.
AC	Q9CY64;			
DC	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DE	2500001N03RIK protein.			
GN	BLVR OR 2500001N03RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
NCBI_TaxID=10090;				
[1]				
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/GJ; TISSUE=Embryonic liver;			
RX	MEDLINE=21085560; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,			
RA	Hayashizaki Y.,			
RT	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK010847; BAB27219.1;			
DR	MGI; MGI:88170; BLVRa.			

```

DR InterPro: IPR000683; GFO_IDH_MoCA.
DR Pfam: PF01408; GFO_IDH_MoCA; 1.
SQ SEQUENCE 295 AA; 33524 MW; F2E1682BD77032A4 CRC64;

Query Match 94.6%; Score 35; DB 11; Length 295;
Best Local Similarity 85.7%; Pred. No. 6.5;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKR1XHC 7
    |||||
DB 274 KKR1LHC 280

RESULT 2
Q9DD21 PRELIMINARY; PRT; 303 AA.
AC Q9DD21: 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 0610006A11RIK protein.
GN BLVRA OR 0610006A11RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21083660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staib F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK002231; BAB21950.1;
DR MGD: MGI:88170; BLVra.
DR InterPro: IPR000683; GFO_IDH_MoCA.
DR Pfam: PF01408; GFO_IDH_MoCA; 1.
SQ SEQUENCE 303 AA; 34491 MW; 52D8A3B02E956EB CRC64;

Query Match 94.6%; Score 35; DB 11; Length 303;
Best Local Similarity 85.7%; Pred. No. 6.7;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKR1XHC 7
    |||||
DB 274 KKR1LHC 280

RESULT 3
Q9M149 PRELIMINARY; PRT; 401 AA.
AC Q9M149: 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative-phosphatidylinositol kinase.

```

```

GN Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN 1
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 2
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL161491; CAB80928.1;
DR InterPro: IPR002498; PIP5K.
DR Pfam: PF01504; PIP5K; 1.
DR SMART: SM00330; PIPKC; 1.
KW Kinase.
SQ SEQUENCE 401 AA; 45659 MW; 8A12D10DA2DED4CA CRC64;

Query Match 94.6%; Score 35; DB 10; Length 401;
Best Local Similarity 85.7%; Pred. No. 8.4;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKR1XHC 7
    |||||
DB 353 KKR1EHC 359

RESULT 4
O04613 PRELIMINARY; PRT; 431 AA.
ID O04613
AC O04613: 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE A_IG002N01.9 protein.
GN A_IG002N01.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Scheet P., Maggi L.;
RT "The sequence of A. thaliana IG002N01.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN 2
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Wash-U;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN 3
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Waterston R.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF007269; AAB61030.1;
DR InterPro: IPR002498; PIP5K.
DR Pfam: PF01504; PIP5K; 1.
DR SMART: SM00330; PIPKC; 1.
SQ SEQUENCE 431 AA; 49356 MW; 07A53F23BDD0942B CRC64;

Query Match 94.6%; Score 35; DB 10; Length 431;
Best Local Similarity 85.7%; Pred. No. 8.9;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKR1XHC 7

```

```

Db          383 KKRIEHC 389
|||||
RESULT 5
Q9FHV3      PRELIMINARY;      PRT;      599 AA.
AC Q9PHV3;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Mutator-like transposase-like protein.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Columbia;
RX MEDLINE=99397451; PubMed=10470850;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT P1 and TAC clones.";
RL DNA Res. 6:183-195(1999).
DR EMBL: AB017068; BAB11366.1; -
DR InterPro: IPR006564; Znf_PWZ.
DR SMART: SM00575; Znf_PWZ; 1.
SQ SEQUENCE 599 AA; 69407 MW; 36BAEE2F2A82D717 CRC64;

Query Match          94.6%; Score 35; DB 10; Length 599;
Best Local Similarity 85.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIEHC 7
|||||
Db 563 KKRIEHC 569

RESULT 6
O49511      PRELIMINARY;      PRT;      633 AA.
AC O49511;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE MUDR transposable element - like protein (MUDR transposable element-
DE like protein).
GN F28J12.70 OR AT4G18410.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RP Beran M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL021710; CAAL16721.1; -
DR EMBL: AL161548; CAB78843.1; -
DR InterPro: IPR004332; MUDR.

```

```

DR InterPro: IPR006564; Znf_PWZ.
DR Pfam: PF03108; MUDR; 1.
DR SMART: SM00575; Znf_PWZ; 1.
SQ SEQUENCE 633 AA; 72930 MW; 3FC298BF2218C623 CRC64;

Query Match          94.6%; Score 35; DB 10; Length 633;
Best Local Similarity 85.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIEHC 7
|||||
Db 591 KKRIEHC 597

RESULT 7
Q03949      PRELIMINARY;      PRT;      189 AA.
AC Q03949; Q8XFB8;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Transposase.
GN AL44868 OR ALR0552 OR ALR1853 OR ALL1972 OR ALL2067 OR ALR2773.
OC Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=91358370; PubMed=1653219;
RX Alam J., Vrba J.M., Cai Y., Martin J.A., Weislo L.J., Curtis S.E.;
RT "Characterization of the IS895 family of insertion sequences from the
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL J. Bacteriol. 173:5778-5783(1991).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL: M67475; AAA98138.1; -
DR EMBL: AP003597; BAB76567.1; -
DR EMBL: AP003582; BAB72510.1; -
DR EMBL: AP003587; BAB73552.1; -
DR EMBL: AP003587; BAB73671.1; -
DR EMBL: AP003588; BAB73766.1; -
DR EMBL: AP003590; BAB74472.1; -
DR PIR: A38117; A38117.
KW Complete proteome.
SQ SEQUENCE 189 AA; 21937 MW; B873A342856C2103 CRC64;

Query Match          89.2%; Score 33; DB 16; Length 189;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIEHC 7
|||||
Db 171 KKRIEHC 177

RESULT 8
Q9D6S2      PRELIMINARY;      PRT;      508 AA.
AC Q9D6S2;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE 2'-5' oligoadenylate synthetase-like.
GN OASL2 OR OASL.

```

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Humé D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 -DR EMBL; AK010034; BAB26655.1; -
 DR MGD; MGI:1344390; Oas12.
 DR InterPro; IPR006117; 25A_SYNTH_2.
 DR InterPro; IPR006116; 25A_SYNTH_UB.
 DR InterPro; IPR001201; PAP_25A_core.
 DR InterPro; IPR000626; ubiquitin.
 DR Pfam; PF00240; ubiquitin; 1.
 DR SMART; SM00213; UBG; 1.
 DR PROSITE; PS00833; 25A_SYNTH_2; 1.
 DR PROSITE; PS0152; 25A_SYNTH_3; 1.
 DR PROSITE; PS50053; UBIQUITIN_2; 1.
 SQ SEQUENCE 508 AA; 58767 MW; 9552B4540CC801A0 CRC64;
 Query Match 89.2%; Score 33; DB 11; Length 508;
 Best Local Similarity 71.4%; Pred. No. 28;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIKHC 7
 Db 109 KKRLIHC 115
 RESULT 9
 OQK234 PRELIMINARY; PRT; 508 AA.
 AC OQK234;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DE Hypothetical protein.
 DE Hypothetical protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC034361; AAH34361.1; -
 DR InterPro; IPR006117; 25A_SYNTH_2.
 DR InterPro; IPR006116; 25A_SYNTH_UB.
 DR InterPro; IPR001201; PAP_25A_core.
 DR InterPro; IPR000626; Ubiquitin.
 DR SMART; SM00213; UBG; 1.
 DR PROSITE; PS00833; 25A_SYNTH_2; 1.

DR PROSITE; PS0152; 25A_SYNTH_3; 1.
 DR PROSITE; PS50053; UBIQUITIN_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 508 AA; 58734 MW; 77F8AF3000A0E888 CRC64;
 Query Match 89.2%; Score 33; DB 11; Length 508;
 Best Local Similarity 71.4%; Pred. No. 28;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIKHC 7
 Db 109 KKRLIHC 115
 RESULT 10
 OQK234 PRELIMINARY; PRT; 510 AA.
 AC OQK234;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Putative NADH dehydrogenase (ubiquinone oxidoreductase).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 RT annotation.";
 RL Genome Biol. 0:0-0(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldmann K.;
 RT "Full-length cDNA from Arabidopsis thaliana.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY084663; AAM61225.1; -
 DR InterPro; IPR001327; FAD_Pyr_redox.
 DR Pfam; PF00070; pyr_redox; 1.
 KW Ubiquinone.
 SQ SEQUENCE 510 AA; 56674 MW; EBE60EF5FA8E052A CRC64;
 Query Match 89.2%; Score 33; DB 10; Length 510;
 Best Local Similarity 71.4%; Pred. No. 28;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIKHC 7
 Db 238 KKRLIHC 244
 RESULT 11
 OQK234 PRELIMINARY; PRT; 510 AA.
 AC OQK234;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN At1G07180/F10K1.8
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,

RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
 RA Hayashizaki Y., Shinozaki K.,
 RT "Arabidopsis thaliana full-length cDNA."
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK118982; BAC33558.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 510 AA; 5628 MW; 12BA0AB1A72A56AB CRC64;
 Query Match 89.2%; Score 33; DB 10; Length 510;
 Best Local Similarity 71.4%; Pred. No. 28;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIHXC 7
 |||: ||
 DB 238 KKRLHLC 244

RESULT 12
 Q9LML0
 ID Q9LML0 PRELIMINARY; PRT; 512 AA.
 AC Q9LML0;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE F10K1.11 protein.
 GN F10K1.11.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Liu S.X., Chan A., Yu G., Etgu P., Lee J.M., Lenz C., Pham P.,
 RA Sakano H., Toriumi M., Chung M., Goldsmith A., Liu A., Liu A.,
 RA Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
 RA Conway A., Hansen N., Johnson-Hopson C., Khan S., Kim C., Lam B.,
 RA Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
 RA Ecker J.R., Federspiel N.A., Theologis A.
 RT "The sequence of BAC F10K1 from Arabidopsis thaliana chromosome 1."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC067971; AAF82202.1;
 DR InterPro: IPR001327; FAD_pyr_redox.
 DR Pfam: PF00070; pyr_redox; 1.
 SQ SEQUENCE 512 AA; 56857 MW; 1F63AFA9A1A2C13B CRC64;
 Query Match 89.2%; Score 33; DB 10; Length 512;
 Best Local Similarity 71.4%; Pred. No. 28;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIHXC 7
 |||: ||
 DB 236 KKRLHLC 242

RESULT 13
 Q8H5X6
 ID Q8H5X6 PRELIMINARY; PRT; 562 AA.
 AC Q8H5X6;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Putative NADH dehydrogenase.
 GN QJ1112.E08.13
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
 RT clone:OJ1112.E08."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP003705; BAC15811.1; -
 SQ SEQUENCE 562 AA; 60257 MW; 31874019733E2A19 CRC64;
 Query Match 89.2%; Score 33; DB 10; Length 562;
 Best Local Similarity 71.4%; Pred. No. 31;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KKRIHXC 7
 |||: ||
 DB 288 KKRLHLC 294

RESULT 14
 Q9VZM4
 ID Q9VZM4 PRELIMINARY; PRT; 109 AA.
 AC Q9VZM4;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE CGI2029 protein.
 GN CGI2029.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Balwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Weissman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003478; AAF47785.1; -
 DR HSSP: P08047; 1SP2.

DR FlyBase; FBgn0035454; CG12029.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zf-C2H2_3.
 DR ProDom; PD000003; Znf_C2H2; 1.
 DR SMART; SM00335; Znf_C2H2; 3.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
 KW Metal-binding; Zinc; Zinc-finger.
 SQ SEQUENCE 109 AA; 13185 MW; 31791156DA22D2BC CRC64;

Query Match 86.5%; Score 32; DB 5; Length 109;
 Best Local Similarity 71.4%; Pred. No. 13;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIHC 7
 :||| ||
 Db 18 KRIHC 24

RESULT 15

Q9N091 PRELIMINARY; PRT; 683 AA.
 AC Q9N091;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Unnamed protein product.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 libraries."
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB046040; BAB01622.1; -
 DR HSSP; P29350; IGWZ.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; Tyr_PP.
 DR Pfam; PF00102; Y_phosphatase; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
 KW Hydrolase.
 SQ SEQUENCE 683 AA; 76905 MW; 14F0C7F8D78DFB9C CRC64;

Query Match 86.5%; Score 32; DB 6; Length 683;
 Best Local Similarity 71.4%; Pred. No. 60;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKRIHC 7
 :||| ||
 Db 376 KRIHC 382

Search completed: July 29, 2003, 09:56:25
 Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2003, 09:51:58 ; Search time 38.9333 seconds
(without alignments)
32.615 Million cell updates/sec

Title: US-09-606-129A-17
Perfect score: 28
Sequence: 1 QKXCXXXK 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues 1107863
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	23	82.1	182	AAU05770	Grape invertase in
2	22	78.6	61	AAO09412	Human polypeptide
3	22	78.6	74	RAM80784	Human haematologic
4	22	78.6	74	RAM80932	Human haematologic
5	22	78.6	74	RAM82111	Human haematologic
6	22	78.6	121	AAK74003	Human colon cancer
7	22	78.6	149	ABP29211	Streptococcus poly
8	22	78.6	195	AAAB23270	Balanus amphitrite
9	22	78.6	293	ABG16619	Novel human diagno

10	22	78.6	294	AAU34112	Staphylococcus aur
11	22	78.6	322	AAU36866	Staphylococcus aur
12	22	78.6	361	AAU34698	Human secreted pro
13	22	78.6	384	ABG15335	Novel human diagno
14	22	78.6	425	ABG06903	Novel human diagno
15	22	78.6	541	AAU93183	Human polypeptide,
16	22	78.6	607	ABG2173	Drosophila melanog
17	22	78.6	617	AAU16310	Human novel secret
18	22	78.6	617	AAU42349	Human polypeptide
19	22	78.6	617	AAU73875	Human colon cancer
20	22	78.6	617	ABU55379	Human novel polype
21	22	78.6	744	ABU56426	Drosophila melanog
22	22	78.6	804	AAU91287	Human NOV5f protei
23	22	78.6	825	AAU91284	Human NOV5c protei
24	22	78.6	1472	AAU31578	Novel human secret
25	21	75.0	62	ABU50864	Helicobacter pylori
26	21	75.0	73	ABP03316	Human ORFX protein
27	21	75.0	79	AAU53799	Human colon cancer
28	21	75.0	86	AAW20389	H. pylori secreted
29	21	75.0	87	AAW20941	H. pylori secreted
30	21	75.0	95	ABP34858	Human ORF3831 prot
31	21	75.0	104	ABG55582	Human breast speci
32	21	75.0	115	ABG14374	Novel human diagno
33	21	75.0	115	AAE03287	Helicobacter pylori
34	21	75.0	128	AAO01272	Human polypeptide
35	21	75.0	135	AAU42571	Human ORFX ORF2335
36	21	75.0	138	ABP41811	Human ovarian anti
37	21	75.0	141	ABU57108	Human prostate can
38	21	75.0	154	ABP57882	Ribosomal protein
39	21	75.0	170	AAU28035	Novel human secret
40	21	75.0	176	AAU08453	Arabidopsis thalia
41	21	75.0	177	AAU28036	Novel human secret
42	21	75.0	178	ABU57063	Mouse ischaemic co
43	21	75.0	183	ABP00866	Human ORFX protein
44	21	75.0	188	AAU15906	Human novel secret
45	21	75.0	188	ABU54975	Human novel polype

ALIGNMENTS

RESULT 1
AAU05770
ID AAU05770 standard; Protein: 182 AA.
XX
AC AAU05770;
XX
DT 24-OCT-2001 (first entry)
XX
DE Grape invertase inhibitor #3.
XX
KW Grape; invertase inhibitor; gene therapy; environmental stress;
KW kernel development; antisense.
XX
OS Vitis sp.
XX
PN WO200158939-A2.
XX
PD 16-AUG-2001.
XX
PF 12-FEB-2001; 2001WO-US04492.
XX
PR 10-FEB-2000; 2000US-0181509.
XX
PA (PTON-) PIONEER HI-BRED INT INC.
PA (DUPO) DU PONT DE NEMOURS & CO E. I.
PI Helentjaris T, Bate NJ, Allen SM;
XX
DR WPI: 2001-502706/55.
DR N-PSDB; AAS11355.
XX
PT An isolated polypeptide when recombinantly expressed in a plant is

PT useful for modulating invertase activity and increasing yield in the
 PT plant
 XX
 PS Claim 1; Page 52; 83pp; English.
 XX The invention relates to novel plant invertase inhibitors or a
 CC yeast invertase which, when recombinantly expressed in a plant, can
 CC modulate invertase activity and increase yield in the plant. Chimeric
 CC invertase inhibitors are useful for modulating invertase activity and
 CC increasing yield in a plant (especially crop species) when used to
 CC transform the plant and are also useful for modulating kernel development
 CC and protecting plants against the harmful/detrimental effects of stress
 CC and adverse environmental conditions. Yeast invertase is less sensitive
 CC to invertase inhibitors therefore is an attractive option to supplement
 CC invertase activity in a plant using gene therapy. The novel
 CC invertase inhibitor nucleic acids may be used in their antisense form.
 CC The present sequence represents a grape invertase inhibitor.
 XX
 SQ Sequence 182 AA;

Query Match 82.1%; Score 23; DB 22; Length 182;
 Best Local Similarity 50.0%; Pred. No. 7e+02; 4; Indels 0; Gaps 0;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
 |||
 Db 34 QKTCRSTK 41

RESULT 2
 AA009412
 ID AA009412 standard; Protein; 61 AA.

AC AA009412;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 23304.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

XX WO200164835-A2.

PN 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AA189343.

XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -

XX Claim 20; SEQ ID NO 23304; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating.
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 61 AA;

Query Match 78.6%; Score 22; DB 22; Length 61;
 Best Local Similarity 50.0%; Pred. No. 5.1e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QKXCXXXK 8
 |||
 Db 32 QKSCSFLK 39

RESULT 3
 AAM80784
 ID AAM80784 standard; Protein; 74 AA.

AC AAM80784;

XX 13-NOV-2001 (first entry)

XX Human haematological malignancy-related antigen #482.

XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;
 KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
 KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

XX Homo sapiens.

XX WO200164886-A2.

XX 07-SEP-2001.

XX 01-MAR-2001; 2001WO-US07272.

XX 01-MAR-2000; 2000US-0186126.

PR 17-MAR-2000; 2000US-0190479.

PR 27-APR-2000; 2000US-0200545.

PR 28-APR-2000; 2000US-0200303.

PR 01-MAY-2000; 2000US-0200779.

PR 04-MAY-2000; 2000US-0200999.

PR 22-MAY-2000; 2000US-0202084.

PR 14-JUL-2000; 2000US-0206201.

PR 03-AUG-2000; 2000US-0218950.

PR 04-AUG-2000; 2000US-0222903.

PR 07-AUG-2000; 2000US-0223416.

XX (CORI-) CORIXA CORP.

XX Gaiger A, Algate PA, Mannion J;

XX WPI; 2001-514842/56.

XX Compositions and methods for the detection of hematological
 PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
 PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -

XX Claim 1; Page 708; 1252pp; English.

XX The present invention relates to compositions and methods for the
 CC detection, diagnosis and therapy of haematological malignancies. The
 CC present sequence is the protein sequence of a human haematological
 CC malignancy related antigen. The methods of the present invention comprise
 CC detecting the presence of haematological malignancy related antigen(s) in

CC a sample obtained from the patient (an increased level of the
 CC polypeptide, compared to an unaffected individual, is indicative of an
 CC increased risk). Haematological malignancies which can be treated using
 CC the present invention are chronic lymphocytic leukaemia, lymphoma,
 CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
 CC cell non-Hodgkin's lymphoma.

XX Sequence 74 AA;

Query Match 78.6%; Score 22; DB 22; Length 74;
 Best Local Similarity 50.0%; Pred. No. 6e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 QKXCXXXK 8
 |||||
 Db 38 QKMCSLSK 45

RESULT 4
 AAM80932
 ID AAM80932 standard; Protein; 74 AA.

XX AC AAM80932;

XX DT 13-NOV-2001 (first entry)

XX Human haematological malignancy-related antigen #630.

XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;

KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
 KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

XX OS Homo sapiens.

XX PN WO200164886-A2.

XX PD 07-SEP-2001.

XX PF 01-MAR-2001; 2001WO-US07272.

XX PR 01-MAR-2000; 2000US-0186126.

XX PR 17-MAR-2000; 2000US-0190479.

XX PR 27-APR-2000; 2000US-0200545.

XX PR 28-APR-2000; 2000US-0200303.

XX PR 28-APR-2000; 2000US-0200779.

XX PR 01-MAY-2000; 2000US-0200999.

XX PR 04-MAY-2000; 2000US-0202084.

XX PR 22-MAY-2000; 2000US-0218950.

XX PR 14-JUL-2000; 2000US-0222903.

XX PR 03-AUG-2000; 2000US-0223416.

XX PR 07-AUG-2000; 2000US-0223378.

XX PA (CORI-) CORIXA CORP.

XX PI Gaiger A, Algate PA, Mannion J;

XX DR WPI; 2001-514842/56.

XX Compositions and methods for the detection of hematological
 PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
 PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
 XX Claim 1; Pages 766-767; 1252pp; English.

XX The present invention relates to compositions and methods for the
 CC detection, diagnosis and therapy of haematological malignancies. The
 CC present sequence is the protein sequence of a human haematological
 CC malignancy related antigen. The methods of the present invention comprise
 CC detecting the presence of haematological malignancy related antigen(s) in
 CC a sample obtained from the patient (an increased level of the
 CC polypeptide, compared to an unaffected individual, is indicative of an
 CC increased risk). Haematological malignancies which can be treated using

CC the present invention are Chronic lymphocytic leukaemia, lymphoma,
 CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
 CC cell non-Hodgkin's lymphoma.

XX Sequence 74 AA;

Query Match 78.6%; Score 22; DB 22; Length 74;
 Best Local Similarity 50.0%; Pred. No. 6e+02;
 Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 QKXCXXXK 8
 |||||
 Db 38 QKMCSLSK 45

RESULT 5
 AAM82111
 ID AAM82111 standard; Protein; 74 AA.

XX AC AAM82111;

XX DT 13-NOV-2001 (first entry)

XX Human haematological malignancy-related antigen #1809.

XX Human; cytostatic; vascular; gene therapy; vaccine; lymphoma;

KW haematological malignancy; antigen; chronic lymphocytic leukaemia;
 KW follicular lymphoma; Hodgkin's lymphoma; non-Hodgkin's lymphoma.

XX OS Homo sapiens.

XX PN WO200164886-A2.

XX PD 07-SEP-2001.

XX PF 01-MAR-2001; 2001WO-US07272.

XX PR 01-MAR-2000; 2000US-0186126.

XX PR 17-MAR-2000; 2000US-0190479.

XX PR 27-APR-2000; 2000US-0200545.

XX PR 28-APR-2000; 2000US-0200303.

XX PR 28-APR-2000; 2000US-0200779.

XX PR 01-MAY-2000; 2000US-0200999.

XX PR 04-MAY-2000; 2000US-0202084.

XX PR 22-MAY-2000; 2000US-0206201.

XX PR 14-JUL-2000; 2000US-0218950.

XX PR 03-AUG-2000; 2000US-0222903.

XX PR 04-AUG-2000; 2000US-0223416.

XX PR 07-AUG-2000; 2000US-0223378.

XX PA (CORI-) CORIXA CORP.

XX PI Gaiger A, Algate PA, Mannion J;

XX DR WPI; 2001-514842/56.

XX Compositions and methods for the detection of hematological
 PT malignancies, e.g. chronic lymphocytic leukemia, lymphoma, follicular
 PT lymphoma and Hodgkin's and T/B cell non-Hodgkin's lymphoma -
 XX Claim 1; Pages 1230-1231; 1252pp; English.

XX The present invention relates to compositions and methods for the
 CC detection, diagnosis and therapy of haematological malignancies. The
 CC present sequence is the protein sequence of a human haematological
 CC malignancy related antigen. The methods of the present invention comprise
 CC detecting the presence of haematological malignancy related antigen(s) in
 CC a sample obtained from the patient (an increased level of the
 CC polypeptide, compared to an unaffected individual, is indicative of an
 CC increased risk). Haematological malignancies which can be treated using
 CC the present invention are chronic lymphocytic leukaemia, lymphoma,
 CC follicular lymphoma, Hodgkin's lymphoma, non-Hodgkin's lymphoma and T/B
 CC cell non-Hodgkin's lymphoma.

```
XX SQ Sequence 74 AA;
Query Match 78.6%; Score 22; DB 22; Length 74;
Best Local Similarity 50.0%; Pred. No. 6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
DB 69 QKACTWAK 76
II I I

RESULT 7
ABP29211
ID ABP29211 standard; Protein; 149 AA.
XX AC ABP29211;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 7598.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus pyogenes.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB04789.
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI: 2002-352536/38.
XX DR N-PSDB; ABN69842.
XX PT New Streptococcus protein for the treatment or prevention of infection
XX PT or disease caused by Streptococcus bacteria, such as meningitis, and
XX PT for detecting a compound that binds to the protein.
XX PS Claim 1; Page 3899; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN1526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
XX CC used as a vaccine or diagnostic composition. The disease caused by
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be
XX CC used in gene therapy. Antibodies to (I) are used for affinity
XX CC chromatography, immunoassays, and distinguishing/identifying
XX CC Streptococcus proteins.
XX SQ Sequence 149 AA;

Query Match 78.6%; Score 22; DB 23; Length 149;
Best Local Similarity 50.0%; Pred. No. 1e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
DB 69 QKACTWAK 76
II I I

RESULT 6
AAG74003
ID AAG74003 standard; Protein; 121 AA.
XX AC AAG74003;
XX DT 03-SEP-2001 (first entry)
XX DE Human colon cancer antigen protein SEQ ID NO:4767.
XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX KW colorectal carcinoma.
XX OS Homo sapiens.
XX PN WO200122920-A2.
XX PD 05-APR-2001.
XX PF 28-SEP-2000; 2000WO-US26524.
XX PR 29-SEP-1999; 99US-0157137.
XX PR 03-NOV-1999; 99US-0163280.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX DR WPI: 2001-235357/24.
XX DR N-PSDB; AAH33434.
XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX PS Claim 11; Page 6565-6566; 9803pp; English.
XX CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX CC cancer-associated nucleic acid molecules (N) and proteins (P), where
XX CC the proteins are collectively known as colon cancer antigens. The colon
XX CC cancer antigens have cytostatic activity and can be used in gene
XX CC therapy and vaccine production. N and P may be used in the prevention,
XX CC diagnosis and treatment of diseases associated with inappropriate P
XX CC expression. For example, N and P may be used to treat disorders
XX CC associated with decreased expression by rectifying mutations or deletions
XX CC in a patient's genome that affect the activity of P by expressing
XX CC inactive proteins or to supplement the patients own production of P.
XX CC Additionally, N may be used to produce the colon cancer-associated Ps
XX CC by inserting the nucleic acids into a host cell and culturing the cell
XX CC to express the proteins. N and P can be used in the prevention, diagnosis
XX CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX CC and AAG77789 represent sequences used in the exemplification of the
XX CC present invention.
XX CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX CC missing at time of publication, meaning no sequences are present for
XX CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX SQ Sequence 121 AA;

Query Match 78.6%; Score 22; DB 22; Length 121;
Best Local Similarity 50.0%; Pred. No. 8.7e+02;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Db 135 QKCSAFK 142

RESULT 8
AAB23270
ID AAB23270 standard; Protein: 195 AA.
XX AC AAB23270;
XX DT 02-FEB-2001 (first entry)
XX DE Balanus amphitrite adhesion/metamorphosis-related protein Bcs-4.
XX KW Adhesion/metamorphosis-related protein Bcs-4; barnacle; larva-specific;
KW adhesion inhibition; metamorphosis inhibition; compound screening;
XX antifouling composition.
XX OS Balanus amphitrite.
XX PN JP2000228985-A.
XX PD 22-AUG-2000.
XX PF 09-FEB-1999; 99JP-0031067.
XX PR 09-FEB-1999; 99JP-0031067.
XX PA (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.
XX DR WPI; 2000-649634/63.
XX DR N-PSDB; AAA97604.
XX Novel barnacle kpiris larva-specific adhesion/metamorphosis-related
PT gene useful for identifying inhibitors of barnacle adhesion
XX
PS Claim 4; Page 19; 32pp; Japanese.
XX The invention relates to six larva-specific adhesion/metamorphosis-
CC related genes from the barnacle Balanus amphitrite (cDNAs given in
CC AAA97601-A97606) and to the proteins they encode (AAB23267-B23272).
CC The genes and the proteins can be used for screening for a
CC substance that inhibits the adhesion or metamorphosis of barnacle
CC larvae, which may be useful in antifouling compositions for use in
CC the shipping industry. The present sequence represents the Balanus
CC amphitrite adhesion/metamorphosis-related protein Bcs-4.
XX
SQ Sequence 195 AA;
Query Match 78.6%; Score 22; DB 21; Length 195;
Best Local Similarity 50.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 QKXCXXXK 8
II I I
Db 86 QKTCADIK 93

RESULT 9
ABG16619
ID ABG16619 standard; Protein: 293 AA.
XX AC ABG16619;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #16610.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 27-SEP-2001.

XX 11-OCT-2001.
XX PD 30-MAR-2001; 2001WO-US08631.
XX PF 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX XX (HYSE-) HYSEQ INC.
XX PA Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS80806.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID NO 46978; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 293 AA;
Query Match 78.6%; Score 22; DB 22; Length 293;
Best Local Similarity 50.0%; Pred. No. 1.7e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 QKXCXXXK 8
II I I
Db 205 QKCTNNK 212

RESULT 10
AAU34112
ID AAU34112 standard; Protein: 294 AA.
XX AC AAU34112;
XX DT 13-FEB-2002 (first entry)
XX DE Staphylococcus aureus cellular proliferation protein #388.
XX KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX OS Staphylococcus aureus.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.

```

XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX XX
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX XX
XX DR WPI: 2001-611495/70.
XX DR N-PSDB: AAS51971.
XX XX
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX XX
XX PS Example 3; Seq ID No 5608; 51lpp; English.
XX XX
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence represents an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 294 AA;
Query Match 78.6%; Score 22; DB 22; Length 294;
Best Local Similarity 50.0%; Pred. No. 1.7e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 QKXCXXXK 8
Db 287 QKACQTLK 294
RESULT 11
AAU36866
ID AAU36866 standard; Protein; 322 AA.
XX AC AAU36866;
XX XX
XX DT 14-FEB-2002 (first entry)
XX XX
XX DE Staphylococcus aureus cellular proliferation protein #1036.
XX KW Antisense; prokaryotic cellular proliferation protein;
XX KW antibiotic; antibacterial; drug design.
XX XX
XX OS Staphylococcus aureus.
XX XX WO200170955-A2.
XX PN
XX PD 27-SEP-2001.

```

```

XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX XX
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX XX
XX DR WPI: 2001-611495/70.
XX DR N-PSDB: AAS54725.
XX XX
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX XX
XX PS Example 3; Seq ID No 12459; 51lpp; English.
XX XX
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence represents an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 322 AA;
Query Match 78.6%; Score 22; DB 22; Length 322;
Best Local Similarity 50.0%; Pred. No. 1.8e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 QKXCXXXK 8
Db 301 QKACQTLK 308
RESULT 12
AAB34698
ID AAB34698 standard; Protein; 361 AA.
XX AC AAB34698;
XX XX
XX DT 26-JAN-2001 (first entry)
XX XX
XX DE Human secreted protein encoded by DNA clone vrl 1.
XX KW Secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;
XX KW systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;
XX KW haematopoiesis regulation; tissue regrowth; wound healing; haemophilia;
XX KW Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;
XX KW contraceptive; infection; growth inhibition; hyperproliferative disorder;
XX KW psoriasis.
XX XX
XX OS Homo sapiens.

```

XX WO20005375-A1.
XX 21-SEP-2000.
XX 17-MAR-2000: 2000WO-US07285.
XX 17-MAR-1999: 99US-0124808.
XX 17-MAR-1999: 99US-0124916.
XX 17-AUG-1999: 99US-0149639.
XX 01-OCT-1999: 99US-0157247.
XX 29-NOV-1999: 99US-0167824.
XX 15-FEB-2000: 2000US-0182711.
XX (ALPH-) ALPHAGENE INC.
XX Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
XX WPI: 2000-638211/61.
XX N-PSDB: AAC59799.
XX Novel proteins and polypeptides useful for the treatment of e.g
XX multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
XX cancer, Alzheimer's disease, Parkinson's disease, stroke, anemia and
XX ulcers
XX Claim 32: Page 400-401; 493pp; English.
XX This invention relates to 59 human secreted proteins and the nucleotide
XX sequences encoding them. Sequences AAC59788-C59846 and AAB34687-B34745
XX represent the proteins and their encoding nucleotide sequences, and
XX sequences AAB34746-B34771 represent fragments of the proteins. Probes
XX for the DNA sequences are represented by sequences AAC59847-C59596. The
XX proteins exhibit neuroprotective, dermatological, immunosuppressive,
XX antiinflammatory, antianemic, nootropic, antiparkinsonian,
XX cerebroprotective, haemostatic, vulnerrary, cytostatic, antipsoriatic,
XX antibacterial, virucide, and fungicide activity. The proteins and
XX nucleotide sequences are useful as nutritional sources or supplements
XX and in research. The proteins are useful for treating immune deficiency
XX and disorders, which may be genetic or resulting from infections,
XX autoimmune disorders such as multiple sclerosis, systemic lupus
XX erythematosus, rheumatoid arthritis, and for treating myeloid or lymphoid
XX cell deficiencies such as anemias by regulating haematopoiesis. The
XX proteins are also useful in compositions for bone, cartilage, tendon,
XX ligament and/or nerve tissue growth or regeneration, for wound healing,
XX tissue repair and replacement and in the treatment of wounds, incisions
XX and ulcers. Other uses include in the treatment of central and
XX peripheral nervous system and neuropathies such as Alzheimer's and
XX Parkinson's diseases and Shy-brager syndrome, and mechanical and
XX traumatic disorders, such as spinal cord disorders, head trauma and
XX stroke. The proteins may also be used as a contraceptive, and for
XX treating coagulation disorders such as haemophilias. The protein and
XX nucleotide sequences with cadherin activity are useful for treating
XX cancer. Other uses for the protein include for inhibiting the growth,
XX infection or function of, or killing, infectious agents such as bacteria,
XX virus, fungi and other parasites, for effecting bodily characteristics
XX such as height, weight, hair colour, effecting biorhythms or cardiac
XX cycles or rhythms, effecting metabolism, catabolism, anabolism, lipid
XX processing, utilization, storage or elimination of dietary fat, lipid,
XX protein, carbohydrate, vitamins, minerals, cofactors, effecting
XX behavioural characteristics, providing analgesic effects and for treating
XX hyperproliferative disorders such as psoriasis.

XX Sequence 361 AA;
Query Match 78.6%; Score 22; DB 21; Length 361;
Best Local Similarity 50.0%; Pred. No. 2e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 OKXCXXXK 8
DB 273 QKCLASK 280

RESULT 13

ABG15335
XX ID ABG15335 standard; Protein; 384 AA.
XX AC ABG15335;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #15326.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001: 2001WO-US08631.
XX 31-MAR-2000: 2000US-0540217.
XX 23-AUG-2000: 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX N-PSDB: AAS79522.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity
XX Claim 20: SEQ ID No 45694: 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 384 AA;

Query Match 78.6%; Score 22; DB 22; Length 384;
Best Local Similarity 50.0%; Pred. No. 2.1e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 OKXCXXXK 8
DB 245 QKCLASK 252

RESULT 14

ABG06903
ID ABG06903 standard; Protein; 425 AA.

XX AC ABG06903;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #6894.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX PF 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX PA Drmanac RT, Liu C, Tang YT;
XX PI WPI: 2001-639362/73.
XX DR N-PSDB; AAS71090.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity
XX PS Claim 20; SEQ ID NO 37262; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG03077 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 425 AA;

Query Match 78.6%; Score 22; DB 22; Length 425;
Best Local Similarity 50.0%; Pred. No. 2.3e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
II I I
Db 337 QKSCNNK 344

RESULT 15

AAM93183
ID AAM93183 standard; Protein; 541 AA.

XX AC AAM93183;

XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide, SEQ ID NO: 2550.
XX KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX OS Homo sapiens.
XX PN EP1130094-A2.
XX PD 05-SEP-2001.
XX PF 07-JUL-2000; 2000EP-0114089.
XX PR 08-JUL-1999; 99JP-0194486.
XX PR 11-JAN-2000; 2000JP-0118774.
XX PR 02-MAY-2000; 2000JP-0183765.
XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX DR WPI: 2001-524255/58.
XX DR N-PSDB; AAK94088.
XX PT 830 Primers useful for synthesizing full length cDNA clones and their
XX PT use in genetic manipulation
XX PS Disclosure; SEQ ID NO 2550; 1380pp + sequence listing; English.
XX CC The invention relates to primers for synthesising full length cDNA
XX CC clones. 830 cDNA molecules encoding a human protein have been
XX CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX CC molecules have been determined. Primers for synthesising the full length
XX CC cDNA are useful for clarifying the function of the protein encoded by
XX CC the cDNA. The full length clones were obtained by construction of full
XX CC length enriched cDNA libraries that were synthesised by the oligo-capping
XX CC method. The primers enable the production of the full length cDNA easily
XX CC without any special methods. The present sequence is a polypeptide
XX CC provided in the specification.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in CD-ROM format directly from EPO.

XX SQ Sequence 541 AA;

Query Match 78.6%; Score 22; DB 22; Length 541;
Best Local Similarity 50.0%; Pred. No. 2.7e+03;
Matches 4; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKXCXXXK 8
II I I
Db 453 QKDCLASK 460

Search completed: July 29, 2003, 09:54:57
Job time : 40.9333 secs